Ville Korpela, Michele Lombardi, and Julius Zachariassen

Behavioral Implementation Without Unanimity

Aboa Centre for Economics
Discussion paper No. 165
Turku
January 2024

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.

ISSN 1796-3133

Printed in Uniprint
Turku
January 2024

Ville Korpela, Michele Lombardi, and Julius Zachariassen

Behavioral Implementation Without Unanimity

Aboa Centre for Economics
Discussion paper No. 165
January 2024

ABSTRACT

Behavioral implementation studies implementation when agents' choices need not be rational. All existing papers of this literature, however, fail to handle a large class of choice behaviors because they rely on a well-known condition called Unanimity. This condition says, roughly speaking, that if all agents would select the same outcome form the set of all feasible outcomes, then this outcome should be deemed socially optimal. While Unanimity is both sensible as a property of a goal and necessary for implementation under rational behavior, with non-rational behavior it is neither. In this paper we investigate behavioral implementation under complete information without assuming Unanimity. Moreover, we give a full characterization of behaviorally implementable SCRs when the designer can use individually based rights structure.

JEL Classification: C72, D11, D71, D82

Keywords: Behavioral economics, Implementation theory, Rights structure, Unanimity

Contact information

Ville Korpela
Turku School of Economics, University of Turku, Finland.
Email: vipeko(at)utu.fi

Michele Lombardi
University of Liverpool Management School, Liverpool, UK.
Email: michele.lombardi(at)liverpool.ac.uk

Julius Zachariassen
Turku School of Economics, University of Turku, Finland.
Email: julius.v.zachariassen(at)utu.fi.

1 Introduction

To organize tasks efficiently the designer of social institutions needs information. Often this information is scattered among different agents and the designer cannot trust that it will be truthfully revealed. The designer must thus rely on the reports of different agents to guide his selection of the outcome. Taking into account that agents might not be truthful, what are the attainable goals of the designer? We study this classic question of mechanism design under complete information. However, instead of assuming that all agents are rational, in the sense of maximizing a context-independent preference relation, we make no assumptions whatsoever about the possible choice behavior. In the following example we illustrate the scope of behaviors this paper analyses.

Decoy effects[1]. A board with three members is tasked with choosing a new colleague. They consider four candidates $\{a, b, a^*, b^*\}$ of whom the first two are considered competent and the last two not. Additionally, candidate a^* is similar to a but dissimilar to b and vice versa. Some members of the board may be sensitive to the decoy effect, i.e., inferior alternatives can influence their decision making. When faced with $\{a, b\}$ they choose their favorite candidate, but when faced with $\{a, a^*, b\}$ they opt for candidate a, even if b was their true favorite. This kind of choice behavior clearly violates standard assumptions of rationality and thus falls beyond the scope of classic mechanism design.

Two seminal papers on behavioral implementation are Korpela

[1] See, e.g., Huber and Pluto (1983) and Herne (1997)

(2012) and de Clippel (2014). Both use the same generalization of Nash equilibrium to a choice function environment.[2] Unfortunately, both papers fail to handle a large class of interesting choice behaviors, like the decoy effect example above, and pretty much for the same reason too; because they assume a property known as Unanimity.[3] In his main characterization results de Clippel (2014) assumes Unanimity or a property called No-Veto Power, which implies Unanimity, while Korpela (2012) assumes that the choice behavior satisfies a property called Sen's α. This property says, roughly speaking, that if an outcome is selected from a set, then it must be selected from any subset that includes it. Under rationality, both of these properties imply that an outcome which is selected from the set of all feasible outcomes by all agents, must be selected from any row of any mechanism that contains it by all. However, most of the interesting behavioral biases that lead agents to select some outcome from the set of all feasible outcomes, do not imply that the same outcome would be chosen from smaller sets that include it, i.e., Unanimity and Sen's α do not hold. This is the case with the decoy effect example above too.

The extent of this problem is made evident by recent progress in behavioral economics. Many experimentally verified non-rational choice behaviors (or anomalies), like limited consideration (Llerasy et al.,2017), choice overload (Toffler, 1970; Settle and Golden

[2]Hurwicz (1986) was the first to generalize Nash equilibrium into a choice function environment. His equilibrium concept, however, does not work for any choice behavior, unlike that of Korpela (2012) and de Clippel (2014).

[3]Under rational behavior Unanimity is both a necessary condition for implementation and a good property of the goal. In contrast, under non-rational behavior it is neither.

1974), self-control problems (Baumeister and Tierney 2011; Kılıçgedik 2022), and temptation effects (Lipman and Pesendorfer 2013), to name but few from a rapidly expanding literature, do not in fact satisfy Unanimity; the outcome that all agents would select among all possible options is not what the designer would like to implement. Furthermore, the behavior does not satisfy Sen's α either, since agents can still make unbiased choices from smaller sets. In this paper we solve this problem by not assuming neither Unanimity nor Sen's α. Moreover, we give a full characterization of implementable SCRs when the designer can use individually based rights structures.

The standard way to do mechanism design is to construct a game form such that the equilibrium outcomes coincides with the goal of the designer. Any game form generates a power distribution in the society; it defines who can replace the current outcome (status quo) and with what outcomes it can be replaced with. One possible way to generalize this idea is to use a *rights structure*, first introduced by Sertel (2001), and later used in implementation theory by Koray and Yildiz (2018) (henceforth KY).[4] Formally, a rights structure consists of a code of rights γ, which specifies the power distribution in the society by assigning a family of coalitions $\gamma(s,t)$ to each pair of distinct states (s,t). If a coalition is assigned to a pair of states (s,t), it is entitled to approve a change from s to t. If no coalition is entitled or willing to move from s to any other state, we regard s to be an equilibrium. We will discuss the full technical details of rights structures in section 2. Our motiva-

[4]See also Korpela, Lombardi, and Vartiainen (2020).

tion for using rights structures instead of game forms is inspired by de Clippel (2014). He conjectured that a general condition for implementation would be so intricate that it would have no practical use. However, using rights structures allows us to approach this problem from a slightly more general angle, and to define a general condition for implementation that is actually rather close to de Clippels (2014) condition. This can be attributed to the fact that rights structure is somewhat more general concept than a game form; any game form is a rights structure but not vise versa.

Our characterization is based on the existence of a collection of sets that satisfy a condition we refer as *extended consistency*. This condition is a generalization of de Clippels' (2014) necessary condition known as consistency. The difference is that for any outcome z which is not optimal at any state, or in other words, used only to generate incentives, there must exist state independent frames, possibly different for different agents, from which the agents will never unanimously select z at any state. These frames do not need to be the set of all feasible outcome and this is how the characterization avoids using Unanimity. Since our result is not dependent on Unanimity or Sen's α, as it is a full characterization, we are able to advance the limits of behavioral implementation.

Related literature. Hurwicz (1986) was the first to study implementation in an environment where the choice behavior of decision makers is represented by a choice function instead of maximising some well-defined preference relation. This choice of modelling has allowed the incorporation of a more rich set of behavior, with various aberrant choices and biases. Along with the rise of behavioral economics, the concept of choice function has been used extensively to model behavioral biases. Besides the seminal papers by de Clippel (2014) and Korpela (2012), there has been plenty of research in implementation theory that takes individual choices as primitivies, e.g., Barlo and Dalkıran (2023a) examine behavioral implementation under incomplete information, Hagiwara (2023) who studies behavioral subgame-perfect implementation, and Hayashi et al. (2023) who investigate behavioral implementation when individuals are allowed to form coalitions. Keeping with the behavioral theme, Caspari and Khanna (2022) investigate matching mechanisms that accommodates non-rational choice behavior. Similarly Chambers and Yenmez (2017) study choice rules applied to matching mechanisms. As a final example, Altun et al. (2023) study Nash implementation under complete information, where the planner has no knowledge of the state of the world nor of the state contingent payoffs of the individuals.[5]

This paper is organized as follows. In section 2 we formally define the choice function environment and other important concepts. In section 3 we present the necessary condition of de Clippel (2014), called consistency, and generalize it to a full characterization. Finally, in section 4, we apply our theorem to two different illustrative examples.

[5]This is an incomplete list. For more on behavioral implementation, we recommend to see also Barlo and Dalkıran (2022; 2023b), de Clippel (2022), Saran R. (2011; 2016)

2 The Setup

The environment consists of n agents $N = \{1, \ldots, n\}$, a set of possible types Θ, and a (non-empty) set of outcomes X. Let $\mathcal{X} = \{A \subseteq X |\ A \neq \emptyset\}$ be the collection of all possible non-empty subsets of X. We focus on complete information environments where the true type is common knowledge among agents but unknown to the designer. Agent i's choice rule for type $\theta \in \Theta$ 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 choice set $A \in \mathcal{X}$. Agent 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 $A \in \mathcal{X}$. If this is not the case, then the choice rule is called non-rational. The goal of the designer is represented by a *social choice rule* (SCR) $F : \Theta \to X$ which maps each type $\theta \in \Theta$ to a set of socially acceptable (or optimal) outcomes $F(\theta) \subseteq X$.

In contrast to standard mechanism design setup, the designer is using a rights structure to implement her goal. Formally, to implement F, the designer constructs a *rights structure* $\Gamma = (S, \gamma, h)$, where S is the state space, $h : S \to X$ is the outcome function, and $\gamma : S \times S \to 2^N$ is a code of rights (a possibly empty-valued correspondence).[6] Subsequently, a code of rights specifies, for each pair of distinct states $s, t \in S$, the family of coalitions $\gamma(s,t) \subseteq 2^N$ that is entitled to move from s to t. If $\gamma(s,t) = \emptyset$, then no coalition is entitled to move from s to t. A rights structure is called an *individually based rights structure* if $\gamma(s,t) \subseteq \{\{i\} \mid i \in N\}$ holds

[6] 2^N is the set of all possible subsets of N.

for all $s, t \in S$. That is, if only singleton coalitions have rights. In this paper we will focus on individually based rights structures only. This is because we want to avoid complications related to behavioral coalition formation[7] and to retain a more direct comparability with the results of de Clippel (2014).

For any state $s \in S$ of a rights structure $\Gamma = (S, \gamma, h)$, we can define the opportunity set $O_i(s) \equiv \{h(t) \in A \mid \{i\} \in \gamma(s,t)\} \cup \{h(s)\}$ of agent i. This is the set of outcomes that agent i can induce from s by unilaterally deviating. State s^* is a *behavioral equilibrium* of rights structure Γ at $\theta \in \Theta$ if $h(s^*) \in C_i(O_i(s^*), \theta)$ holds for all $i \in N$. That is, if all agents select outcome $h(s^*)$ from their respective opportunity sets. Let us denote the set of all behavioral equilibria of Γ at θ by $BE(\Gamma, \theta)$. This is a direct generalization of the equilibrium used by de Clippel (2014) and Korpela (2012) to a right structure environment. We say that rights structure Γ behaviorally implements the SCR F if $F(\theta) = h(BE(\Gamma, \theta))$ holds for all $\theta \in \Theta$. If a rights structure like this exists, then we say that F is behaviorally implementable.

3 Main Result: A Full Characterization

In this section we provide a full characterization of behaviorally implementable SCRs. It is based on a condition known as *consistency* which de Clippel (2014) finds to be necessary for extending Nash implementation beyond the rational domain.[8]

[7] See Hayashi et al. (2023).

[8] Don't get confused; we defined two different kinds of opportunity sets. The first one is defined for a state of a given rights structure. The second one, defined next, is simply a set of outcomes not

Definition 1 (de Clippel, 2014; pp. 2982). *A collection* $\mathcal{O} = \{\mathcal{O}_i(a,\theta) \in \mathcal{X} \mid \theta \in \Theta, a \in F(\theta), i \in N\}$ *of sets is* consistent with F *if:*

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

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

Studying implementation in the Nash equilibrium is based on Maskin (1999; circulated since 1977), who proved that any SCR which 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.[9] Suppose that F is behaviorally implementable. If a is a behavioral equilibrium outcome at θ, the equilibrium state $s \in S$ supporting it defines an opportunity set for each agent i, denoted by $O_i(s,\theta)$, which represents the set of outcomes that agent i can induce by unilaterally deviating from s. From the definition of the behavioral equilibrium, each agent i must choose a from $O_i(s,\theta)$ at θ. Moreover, if there is an alternative type θ' such that every agent i chooses a from $O_i(s,\theta)$ at θ', then s forms a behavioral equilibrium also at θ'. Hence, a is still a F optimal outcome at θ' if F is behaviorally implementable. This explains conditions (i) and (ii) of consistency.[10]

We generalize this condition into a full characterization as follows.

related to any particular right structure

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

[10] Although de Clippel (2014) does not use right structures for implementation, his consistency is necessary for implementation even with rights structures (by the same argument).

Definition 2 (Extended Consistency). *A collection* $\mathcal{O} = \{\mathcal{O}_i(a,\theta) \in \mathcal{X} \mid \theta \in \Theta, a \in F(\theta), i \in N\} \cup \{\mathcal{O}_i(b) \mid b \in B \setminus F(\Theta),\ i \in N\}$ *of sets satisfy extended consistency with respect to* F *if there exists a set* $B \subseteq A$ *such that:*

(i) $\mathcal{O}_i(a,\theta) \subseteq B$ *for all* $\theta \in \Theta$, $a \in F(\theta)$, $i \in N$, and $\mathcal{O}_i(b) \subseteq B$ for all $b \in B \setminus F(\Theta)$, $i \in N$.

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

(iii) *For all* $\theta, \theta' \in \Theta$ *and all* $a \in F(\theta)$, *if* $a \in C_i(\mathcal{O}_i(a,\theta),\theta')$ *for all* $i \in N$, *then* $a \in F(\theta')$.

(iv) *For all* $b \in B \setminus F(\Theta)$,[11] and all $\theta \in \Theta$, $b \notin C_i(\mathcal{O}_i(b),\theta)$ holds for at least one $i \in N$.

Theorem 1 (Necessity). If F is behaviorally implementable by a rights structure, then there exists a collection of sets that satisfies extended consistency with respect to F.

Proof. Suppose that rights structure $\Gamma = (S,\gamma,h)$ implements F in behavioral equilibrium. This rights structure allows us to define the sets in the definition of extended consistency. Let $B \equiv h(S)$. For any $\theta \in \Theta$, $a \in F(\theta)$, and $i \in N$, select an equilibrium state $s^* \in BE(\Gamma,\theta)$ for which $h(s^*) = a$ and set $\mathcal{O}_i(a,\theta) \equiv O_i(s^*)$. This is the set of all outcomes that agent i can induce by unilaterally deviating from state s^*. In addition, for all $b \in B \setminus F(\Theta)$ and all $i \in N$, there exists a state $s \in S$, such that $h(s) = b$. A state like this must exist since $b \in B$. Set $\mathcal{O}_i(b) \equiv O_i(s)$.

[11] $F(\Theta)$ is the range of F i.e. the set $\{x \mid x \in F(\theta) \text{ for some } \theta \in \Theta\}$. Therefore, the set $B \setminus F(\Theta)$ includes all those outcome that are never optimal but merely used to create incentives.

Next we show that F satisfies extended consistency with respect to these sets.

By definition these sets are subsets of the range $h(S)$, thus item (i) of extended consistency holds. Since state s^* is a behavioral equilibrium of Γ at $\theta \in \Theta$, we know that $h(s^*) \in C_i(O_i(s^*), \theta)$. By our earlier construction, $h(s^*) = a$ and $\mathcal{O}_i(a, \theta) \equiv O_i(s^*)$, thus $a \in C_i(\mathcal{O}_i(a, \theta), \theta)$ for all $a \in F(\theta)$, all $\theta \in \Theta$ and all $i \in N$. Hence also item (ii) of extended consistency holds. Moreover, if there exists $\theta' \in \Theta$ such that $a \in C_i(\mathcal{O}_i(a, \theta), \theta')$ holds for all $i \in N$, then s^* is a behavioral equilibrium also at θ' by definition, and therefore $a \in F(\theta')$ must hold since Γ implements F. This confirms that item (iii) holds. Finally, outcomes $b \in B \setminus F(\Theta)$ are never chosen by all $i \in N$ for any $\theta \in \Theta$, so there must exist some $j \in N$ such that $b \notin C_j(\mathcal{O}_j(b), \theta)$ for any $\theta \in \Theta$. Hence item (iv) holds as well. □

In the next theorem we show that extended consistency is not only necessary, but also sufficient for behavioral implementation, and therefore a full characterization.

Theorem 2 (Sufficiency). If SCR F satisfies extended consistency with respect to the collection $\mathcal{O} = \{\mathcal{O}_i(a, \theta) \in \mathcal{X} \mid \theta \in \Theta, a \in F(\theta), i \in N\} \cup \{\mathcal{O}_i(b) \mid b \in B \setminus F(\Theta),\ i \in N\}$ of sets, then it is implementable in behavioral equilibrium.

Proof. We need to construct a rights structure Γ that implements F. To this end, let $\psi : F(\Theta) \to \Theta$ be any one-to-one function such that $\psi^{-1}(\theta) \in F(\theta)$ for all $\theta \in \Theta$. This function connects all outcomes $a \in F(\Theta)$ to exactly one type $\psi(a) \in \Theta$ (whichever) for which the outcome is optimal. Let the set of states be

$$S \equiv \{(a,\theta) \mid \theta \in \Theta,\ a \in F(\theta)\} \cup B,$$

and define the outcome function $h : S \to Z$ as $h((a,\theta)) = a$ for all $(a,\theta) \in S$ and $h(b) = b$ for all $b \in B$. Finally, define the code of rights $\gamma : S \times S \to 2^N$ by the following rules:

(1) $\{i\} \in \gamma((a,\theta),b)$ iff $b \in \mathcal{O}_i(a,\theta)$,

(2) for all $a \in F(\Theta)$; $\{i\} \in \gamma(a,b)$ iff $b \in \mathcal{O}_i(a,\psi(a))$,

(3) for all $b \in B \setminus F(\Theta)$; $\{i\} \in \gamma(b,c)$ iff $c \in \mathcal{O}_i(b)$, and

(4) $\gamma(s,s') = \emptyset$ for all other states s and s'.

Next we need to show that this rights structure implements F in behavioral equilibrium. First, for all $\theta \in \Theta$, and all $a \in F(\theta)$, we must identify the existence of a behavioral equilibrium state s^* such that $h(s^*) = a$. Suppose that F satisfies extended consistency with respect to the collection $\mathcal{O}$ as described before. Let us show that (a,θ) is such a state. By construction the opportunity set at state (a,θ) is $O_i((a,\theta)) = \mathcal{O}_i((a,\theta))$ for all $i \in N$ (see rule 1), and thus by item (ii) of extended consistency, all agents choose a from this opportunity set. This implies that (a,θ) is a behavioral equilibrium such that $h((a,\theta)) = a$. Hence $F(\theta) \subseteq h(BE(\Gamma,\theta))$.

Next we show that $h(BE(\Gamma,\theta)) \subseteq F(\theta)$. Let $\theta \in \Theta$ be the true type. Suppose that $s^* \in BE(\Gamma,\theta)$. There are three possible and mutually exhaustive cases that we must check: (i) $s^* = (b,\theta')$ for some $\theta' \in \Theta$ and $b \in F(\theta')$, (ii) $s^* \in F(\theta)$, or (iii) $s^* \in B \setminus F(\Theta)$. If $s^* = (b,\theta')$, then opportunity sets are defined by rule (1), and therefore $C_i(\mathcal{O}_i(b,\theta'),\theta)$ must hold for all $i \in N$. Thus, by items (ii) and (iii) of extended consistency it follows that $b \in F(\theta)$.

Suppose, then, that $s^* = a \in F(\Theta)$. Thus, rule (2) applies. Since state a is a behavioral equilibrium it must be that $a \in C_i(\mathcal{O}_i(a,\psi(a)),\theta)$ holds for all $i \in N$. Again, by item (ii) of extended consistency, it must be that $a \in F(\theta)$. As a final case, suppose that $s^* = b \in B \setminus F(\theta)$. Now opportunity sets are defined by rule (3). Thus, $b \in C_i(\mathcal{O}_i(b),\theta)$ must hold for all $i \in N$, which is a contradiction with item (iv) of extended consistency. Therefore, states $B \setminus F(\Theta)$ can never be equilibrium states.

By this we have proved that also $h(BE(\Gamma,\theta)) \subseteq F(\theta)$ holds. Thus, this rights structure implements F. □

Corollary 1 (Full characterization). An SCR F is implementable in behavioral equilibrium if, and only if, it satisfies extended consistency with respect to some collection of sets.

Proof. Follows directly from Theorems 1 and 2. □

It is natural to ask whether consistency and extended consistency can sometimes coincide.

Corollary 2 (Equivalence). If $F(\Theta) = X$, then the existence of a consistent collection of sets for F is a sufficient condition for behavioral implementation in a rights structure environment.

Proof. We have shown that extended consistency is a sufficient condition for behavioral implementation in Theorem 2. If $F(\Theta) = X$, then extended consistency cannot hold unless we select $B = X$. But then items (i) and (iv) of extended consistency hold vacuously.

Thus extended consistency coincides with items (ii) and (iii) i.e. with consistency. □

4 Applications

To illustrate extended consistency we study two different application that cannot be handled with any of the results given in de Clippel (2014).

4.1 Limited willpower

Consider a group of n agents trying to achieve a common long-term goal. These agents face tempting outcomes, that they may prefer over the common long-term goal in the short-term. Every agent can exercise some limited willpower[12], which is characterized by the number of tempting outcomes an agent can ignore over their long-term goal. Agents long-term preferences are represented by an ordering $\succ_L$ on X and short-term cravings by $\succ_S$ on X. Let an integer k_i determine i's willpower. Agent i's choice out of any set $A \subseteq X$ is the most preferred outcome with respect to $\succ_L$ among outcomes that are dominated by at most k_i outcomes according to $\succ_S$. The following illustrative example was given by de Clippel (2014); if the willpower of i is k_i=1 and long-term goal and cravings are such that, salad $\succ_L$ pizza $\succ_L$ burger and burger $\succ_S$ pizza $\succ_S$ salad, the agent i will choose pizza from the set {salad, burger, pizza}.

Suppose now that a type $\theta \in \Theta$ determines the common long-term goal and agents' possibly distinct short-term preferences. We de-

[12] For more on willpower, see Kılıçgedik (2022).

fine the SCR F as a function that chooses the common long-term goal from the set of all outcomes for any $\theta \in \Theta$. It is clear that this SCR does not satisfy Unanimity. When agents are faced with the set of all possible outcomes that is large enough, they cannot exercise willpower, which leads them to choose the tempting short-term outcome. If everyone chooses the same short-term outcome, which is a possibility, then that outcome should be selected by F. Thus this SCR does not satisfy Unanimity nor Sen's α. In fact, this SCR doesn't satisfy either of de Clippel's sufficient conditions.[13] We will now show that the SCR described above satisfies extended consistency if $\sum_{i \in N} k_i \geq |X|$.[14]

Since this example assumes that all long-term preferences are possible we have $F(\Theta) = X$. Thus consistency is a sufficient condition by Corollary 2. For all $\theta \in \Theta$, all $a \in F(\theta)$, and all $i \in N$, let $\mathcal{O}_i(a, \theta)$ be such that $|\mathcal{O}_i(a, \theta)| = k_i + 1$, $a \in \mathcal{O}_i(a, \theta)$, and $X \subseteq \bigcup_{i \in N} \mathcal{O}_i(a, \theta)$. We know that these sets exist by the assumption made above.

We only need to check items (ii) and (iii) of extended consistency. Since every $\mathcal{O}_i(a, \theta)$ contains $k_i + 1$ outcomes, any choice is dominated by at most k_i outcomes, thus the most preferred outcome will be always chosen. This implies that the sets satisfy (ii) of extended consistency. Now consider any two types $\theta, \theta' \in \Theta$. For all $a \in F(\theta)$, if $a \in C_i(\mathcal{O}_i(a, \theta), \theta')$ holds for all $i \in N$, then $F(\theta') = \{a\}$, since the most preferred long-term outcome will be chosen by ev-

[13]de Clippel (2014) calls these condition 2B and condition 2B'. He shows that this SCR is implementable by constructing the implementing game form because his theorems cannot handle the example.

[14]Also de Clippel assumes this.

eryone, and it must therefore be a. Hence also (iii) is satisfied.

4.2 Decoy Effects

Let us continue with the decoy example from introduction to further illustrate our results. The decoy effect is known as a phenomenon where agents' choice behavior can be misguided simply due to the existence of so-called decoy outcomes. Suppose a board of three members has to make a decision on hiring a new job candidate. The set of candidates is $\{a, b, a^*, b^*\}$ where the first two are considered by all board members competent and the last two are not. Let candidate a^* be similar to a but dissimilar to b and vice versa. The board members can also choose to hire none of the candidates; an outcome denoted by nh. The preferences of those board members that are able to evaluate the candidates are either $a \succ nh \succ b$ or $b \succ nh \succ a$. We'll refer to board members endowed with these preferences respectively as a-type and b-type. Candidates a^* and b^* are considered worst by all. It is possible, however, that a board member is not able to evaluate the candidates and has preferences $a \sim n \sim b$. These members are sensitive to the decoy effect, so they'll choose candidate x whenever x^* is also available. When faced with a double decoy effect, i.e., a set with $a, a^*, b,$ and b^*, the board member chooses both a and b. We'll refer to these board members as biased types. Let the SCR F be a function that selects the most preferred outcome of the majority as long as the majority does not consist of biased types. Otherwise chooses nh.

Let us construct the sets required in extended consistency. Set

$B = \{a, a^*, b, b^*, nh\}$. For any $\theta \in \Theta$, if $F(\theta) = \{a\}$, then the majority must consist of a-types. Then, set $\mathcal{O}_i(a,\theta) = \{a, b, b^*, nh\}$ for all board members i in this majority, and $\{a\}$ for the rest. If for some other $\theta' \in \Theta$, we have $F(\theta') = \{b\}$, it must be that the majority consists of b-types, in which case set $\mathcal{O}_i(b,\theta') = \{b, a, a^*, nh\}$ for all board members i in this majority, and $\{b\}$ for the rest. Lastly, if for some $\theta'' \in \Theta$ we have $F(\theta'') = \{nh\}$, either there is a majority of biased types or there is no majority. In either case, set $\mathcal{O}_i(nh,\theta'') = \{a, b, nh\}$ for the biased types and $\{x, nh\}$ for the rest, where $x = a$, if the board member is b-type and vice versa. Finally, for $a^*, b^* \in B \setminus F(\Theta)$, set $\mathcal{O}_i(a^*) = \{a^*, a, nh\}$ and $\mathcal{O}_i(b^*) = \{b^*, b, nh\}$ for all agents i.

Consider first the case $F(\theta) = \{a\}$, for some $\theta \in \Theta$. The a-type majority chooses $a \in C_i(\mathcal{O}_i(a,\theta),\theta)$, as well as the rest since there is no other options available. Similar logic holds when, for some $\theta' \in \Theta$, we have $F(\theta') = \{b\}$. Lastly consider some $\theta'' \in \Theta$, for which $F(\theta'') = \{nh\}$. A biased majority will consider all outcomes equally good, so we have $nh \in C_i(\mathcal{O}_i(nh,\theta''),\theta'')$. If the minority board member is a-type, she chooses nh and vice versa for b-type. In the case of no majority, each member still chooses nh from their respective opportunity set. Thus (ii) of extended consistency holds. Next, for any $\theta, \theta' \in \Theta$, if $F(\theta) = \{a\}$, but $F(\theta') \neq a$, we must have a b-type or a biased majority at θ'. At least one board member from this majority must choose from $\{a, b, b^*, nh\}$ and both types clearly choose $\{b\}$. Same holds conversely if $F(\theta) = \{b\}$, but $F(\theta') \neq \{b\}$. If instead $F(\theta) = \{nh\}$, but $F(\theta') \neq \{nh\}$, we must have an a-type or a b-type majority at θ'. If there originally was a biased majority, then at least one member of the new majority must choose from $\{a, b, nh\}$. If instead there was originally no majority, at least one member of the new (e.g., a-type majority) would have to choose from $\{a, b, nh\}$ or $\{a, nh\}$. Similar logic holds for a b-type majority. So (iii) of extended consistency is satisfied. Finally, no one ever chooses a^* or b^* from the sets $\mathcal{O}_i(a^*)$ and $\mathcal{O}_i(b^*)$ respectively. Hence (iv) holds.

Notice that this SCR does not satisfy Unanimity at $X = \{a, a^*, b, b^*, nh\}$. If all agents are biased, candidates a and b would have to be chosen. It would not satisfy consistency either, i.e. items (ii) and (iii), unless a^* and b^* are used to create incentives.

5 Conclusion

We have presented a complete characterization of behaviorally implementable SCRs and also shown when it coincide with consistency. Rather than using classic game forms, we have used the rights structure framwork, formalized by Koray and Yildiz (2018). Our result shows that a generalization of de Clippels (2014) consistency, called extended consistency, is both necessary and sufficient for implementation in a rights structure framework. Also, in a rights structure environment, if $F(\Theta) = X$, then extended consistency and consistency are equivalent.

We hope our contribution paves the way for more research in behavioral mechanism design, especially in choice function environments. An interesting direction for future research would be to consider behavioral interdependence (e.g., herd behavior), since often agents engage in non-rational behavior only when they see others to do the same. While typically herd behavior is viewed as a result of incomplete information, in many instances (e.g., trends) individuals can be seen simply as conforming to the behavior of others. It would be interesting to see if this type of behavior could be modeled using our approach by, e.g., using a suitable domain restriction.

References

[1] Altun OA, Barlo M, Dalkıran NA. Implementation with a sympathizer, Math. Soc. Sciences 121 (2023) 36-49.

[2] Barlo M, Dalkıran NA. Computational implementation. Rev Econ Design (2022). https://doi.org/10.1007/s10058-021-00282-3

[3] Barlo M, Dalkıran NA. Behavioral implementation under incomplete information, Journal of Economic Theory 203 (2023a) 105738.

[4] Barlmo M, Dalkıran NA. Robust behavioral implementation (2023b) Mimeo.

[5] Baumeister R.F, Tierney J. Willpower: rediscovering the greatest human strength (2011) New York: Penguin Press.

[6] Caspari, G. Khanna, M. Non-standard choice in matching markets (2022) ZEW - Center for European Economic Research Discussion Paper No. 22-054, Available at SSRN: https://ssrn.com/abstract=4288333.

[7] Chambers, C.P. Yenmez, M.B. Choice and matching. Am Econ Rev: Microeconomics 9 (2017) 126-147.

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

[9] de Clippel G. Departures from preference maximization, violations of the sure-thing principle, and relevant implications (2022) Mimeo.

[10] Hagiwara, M. Behavioral subgame-perfect implementation (2023) Available at SSRN: https://ssrn.com/abstract=4556468.

[11] Hayashi T, Jain R, Korpela V, Lombardi M. Behavioral strong implementation. *Forthcoming* in Economic Theory, (2023).

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

[13] Huber J, Puto C. Market boundaries and product choice: Illustrating attraction and substitution effects. Journal of Consumer Research 10 (1983) 31-44.

[14] 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.

[15] Kılıçgedik G. Implementation with individuals limited willpower (2022) Master's thesis. SabancıUniversity.

[16] Koray S, Yildiz K. Implementation via rights structures. J. Econ. Theory 176 (2018) 479-502.

[17] Korpela V, Lombardi M, Vartiainen H. Do coalitions matter in designing institutions? J. Econ. Theory 185 (2020).

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

[19] Lipman B.L, Pesendorfer W. Temptation. In Advances in Economics and Econometrics: Tenth World Congress. Vol 1. eds. Acemoglu D, Arellano M, Dekel E. (2013) 243-288. New York: Cambridge University Press

[20] Llerasy J, Masatliogluz Y, Nakajimax D, Ozbay E: When more is less: Limited consideration. Journal of Economic Theory 170 (2017) 70-85.

[21] Maskin E. Nash equilibrium and welfare optimality. Rev Econ Studuies 66 (1999) 23-38.

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

[23] Saran R. Menu-dependent preferences and revelation principle. Journal of Economic Theory 146 (2011) 1712-1720.

[24] Saran R. Bounded depths of rationality and implementation with complete information. Journal of Economic Theory 165 (2016) 517-564.

[25] Sertel R.M. Designing rights: invisible hand theorems, covering and membership. Mimeo. Bogazici University (2001).

[26] Settle R.B, Golden L.L. Consumer perceptions: overchoice in the market place. Proceedings of the 4th Annual Conference of the Association for Consumer Research (1973) 29-37

[27] Toffler A. *Future Shock*. Random House, New York (1970).

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133