Behavioral Implementation[*]

Geoffroy de Clippel[†]

This Version: June 2012

Abstract

Implementation theory assumes that participants' choices are rational, in the sense of being derived from the maximization of a context-independent preference. The paper investigates implementation under complete information when the mechanism designer is aware that individuals suffer from cognitive biases that lead to violations of IIA, or cannot exclude the possibility of such "irrational" behavior.

1 Introduction

Implementation under complete information is a classic problem in mechanism design. The designer would like to implement a rule that selects acceptable outcomes as a function of a problem's characteristics. Unfortunately, while commonly known among participants, these characteristics are unknown to him. He must thus rely on their reports to tailor his selection of outcomes. Taking into account the participants' incentives to misrepresent their information, what are the rules that the designer can effectively implement?

Characteristics encode participants' preferences in standard implementation models. However, there is ample evidence in marketing, psychology and behavioral economics that people's choices need not be consistent with the maximization of a preference relation. Classic examples, which have played

a key role in recent developments in choice theory,[1] include status-quo biases, attraction, compromise and framing effects, temptation and self-control, consideration sets, and choice overload. This paper expands implementation theory so as to be applicable in these circumstances as well. States will encode the participants' choice functions instead of their preferences, and results will be derived without requiring these choice functions to be rationalizable.

To illustrate the scope of my analysis, here are a few out of many applications that fit my framework, but not standard models.

(a) (Hiring with Attraction Effect) Members of a hiring committee are meeting to select a new colleague. Up to six candidates are considered: a, b, c, a', b' and c'. The first three candidates are above the bar, while the last three fall below. For each $x \in \{a, b, c\}$, x' is similar to x, but dissimilar to other candidates. In the spirit of the "attraction" effect,[2] the committee members' individual choices may be influenced by the availability of a similar inferior alternative, e.g. choosing b out of $\{a, b, c\}$, but choosing a out of $\{a, a', b, c\}$. Finding which outcomes can be selected as a function of committee members' characteristics thus falls beyond the scope of standard implementation theory.

(b) (Matching with Consideration Sets) Mechanism design has been applied extensively to matching. Shapley and Scarf's (1974) house allocation problem is one of the simplest matching environments, and more generally one of the simplest examples of exchange economy. If individuals have an underlying preference, but pick options out of feasible sets by maximizing that ordering over a subset of options they actively consider (e.g. m houses that are closest to their workplace, for some fixed integer m), then choices may appear irrational. This behavior falls in the general category of choice with consider-

---

[1]See e.g. Kalai et al. (2002), Manzini and Mariotti (2007), Ambrus and Rozen (2009), and de Clippel and Eliaz (2012) for choices resulting from the combination of multiple conflicting selves, see Masatlioglu and Ok (2005) for choices with a status-quo bias, see Rubinstein and Salant (2006) for choices with order effects, see Bernheim and Rangel (2009) and Salant and Rubinstein (2008) on framing, see Cherepanov et al. (2009), Manzini and Mariotti (2012), Lleras et al. (2010), or Masatlioglu et al. (2012) on limited attention, see Lipman and Pesendorfer (2011) for a survey on choices with temptation and self-control.

[2]First identified by Huber et al. (1982), it has been documented since then in various empirical and experimental settings, see references in Ok et al. (2011, footnotes 2 and 3).

ation sets that has been documented in the marketing literature (see Wright and Barbour (1977)) and studied more recently in the economics literature as well (see Cherepanov et al. (2009), Lleras et al. (2010), Manzini and Mariotti (2012), or Masatlioglu et al. (2012)).

(c) (Collective Choice with Limited Willpower) Individuals in a support group are committing to make joint decisions. They take part in this group to achieve a common long-term goal. The problem is that individual choices are also influenced by a conflicting short-term goal. In a stylized model, suppose for instance that individuals on their own pick the option that is best for the long-term goal, under the constraint that there are at most k superior alternatives for the short-term goal. The parameter k thus captures the individual's willpower in this example. Violations of the property of independence of irrelevant alternatives (IIA), a classic necessary and sufficient condition for rationality, are then likely to occur: one may be able to resist eating a slice of pizza for lunch when the alternative is a salad, but unable to resist both the burger and the pizza slice, and go for the slice if these two options are available on the menu in addition to the salad.[3]

(d) (Groups as Participants)[4] The president of a University is consulting the chairs of its various departments to implement a new policy. The chair of each department knows the preferences of its members, and make choices given what is feasible by following some aggregation procedure. As is well documented in the social choice literature, the choice behavior of the chairs will (in fact, must) violate IIA in most cases.

A social choice rule (SCR) associates a set of outcomes to each state. A mechanism is a collection of sets of messages, one such set for each individual, and a function that associates an outcome to each profile of messages. A mechanism implements a SCR if the set of equilibrium outcomes coincides with the set of outcomes prescribed by the SCR, at every state. Implementation

---

[3]Such choice pattern is consistent with $k = 1$, a long-term goal that ranks the salad above the pizza slice, in turn above the burger, and an opposite short-term goal.

[4]This interpretation motivated Hurwicz's (1986) work, a discussion of which is available at the end of this Introduction.

admits different meanings depending on what is meant by an "equilibrium." I will focus on the two main forms of implementation: in Nash equilibrium and in dominant strategies (see definitions in Section 2).

Section 3 is devoted to the identification of necessary, and sufficient conditions for a SCR to be implementable in either sense, thereby extending classic results by Dasgupta et al. (1979) and Maskin (1999),[5] from the special case of domains containing only rational choice functions to *any* domain of choice functions. Identifying these properties is critical to understand the limits of implementation in applications, some of which will be covered in this paper. Perhaps more insightful than their specific mathematical formulation, it turns out that intuitive analogues of Maskin monotonicity - based either on Bernheim and Rangel's (2009) extended revealed preferences, or directly on choices - are inadequate for characterizing SCRs that are Nash implementable. Coincidentally, the notion of "augmented revelation mechanism," introduced by Mookherjee and Reichelstein's (1990) to characterize Nash-implementability on the rational domain, happens to play a key role in characterizing implementation in dominant strategies in my framework.

Shedding light on the implementation problem requires more than stating the abstract conditions for implementability. Instead, one must understand what they imply in applications. Here are some of the main results I derive.

(i) (Efficiency) The Pareto correspondence is one of the most well-known examples of Nash implementable SCR. An outcome is Pareto efficient if there is no alternative that is preferred by all participants. Rooted in individual preferences, this notion does not immediately apply in the absence of IIA. Bernheim and Rangel (2009) extend the Pareto correspondence to general choice functions: an outcome is BR-efficient if there is no alternative that unambiguously dominates it, where an outcome o' unambiguously dominates an outcome o if o is never picked out of feasible sets that contain o'. I show that BR-efficiency is *not* Nash implementable. On the other hand, I provide in Section 4 an

---

[5]The paper circulated as a working paper from 1977 and 1998. Surveys on the large literature on implementation theory include Maskin (1985), Moore (1992), Palfrey (1992, 2002), Corchón (1996), Jackson (2001), Maskin and Sjöström (2002), and Serrano (2004).

alternative (intuitively appealing) definition of efficiency that coincides with Pareto's definition when individuals are rational, systematically selects a non-empty subset of Bernheim and Rangel's concept, and is Nash implementable on *all* domains (see Proposition 4).

(ii) (Core in Simple Matching Problems) The core of an exchange economy is another important SCR that is Nash implementable. An allocation is in the core if there is no way to make the members of any coalition better off by re-allocating their endowments. As in the case of efficiency, this concept does not apply in the absence of IIA. Section 5 studies the core in the context of the house allocation problem. Bernheim and Rangel's (2009) extended revealed preference leads to a core that is not always Nash implementable. I suggest a sensible alternative definition, and show that it is Nash implementable on *all* domains (see Proposition 5).

(iii) (Rich Domains) The literature contains numerous "impossibility results," which are useful to delineate the limits of implementation. The main lesson from that literature is that dictatorial rules are essentially the only implementable single-valued SCRs if participants may hold many varied preferences. Kalai et al.'s (1979) free triple condition captures the essence of this rich domain assumption. Inspired by it, I propose a choice-based notion of rich domain in the absence of IIA, and use it to extend classic impossibility results (see Proposition 6). Violations of IIA make this extension a bit more permissive though. In the hiring example discussed in (a) above, for instance, there exist implementable SCRs that vary with the reports of participants other than the dictator (see Proposition 7).

(iv) (Specific Classes of Biases) Results descibed so far are general in that they apply to any or numerous classes of individual choice behaviors, with no particular restrictions in mind. The abstract conditions for implementability also prove useful to understand the limits of implementation in the presence of specific biases. Three applications will serve as illustration. Given the prominence of such procedures in the recent choice literature, I start by studying problems involving individuals who maximize preferences over fixed consideration sets. The resulting domain is not rich even if underlying preferences are

unrestricted. Unfortunately, strong negative results apply in this setting as well (see Proposition 8). More permissive results exist when considering other classes of biases. Collective choice problems with limited willpower, as presented in (c) above, offer a first example. A simple mechanism indeed allows to systematically achieve the best element according to the common long-term objective if sufficiently many individuals participate (see Proposition 11). The mechanism effectively allows to add up the participants' limited willpower, resulting in a larger aggregate willpower for the group. Section 7.2 also identifies a class of biases in the context of choice from lists (Rubinstein and Salant (2006)) for which there exists an anonymous single-valued SCR that is both Nash implementable and implementable in dominant strategies.

Choices can sometimes be influenced by external conditions, another form of "irrationality." For instance, the meal picked in a cafeteria may vary with the order in which options are displayed, or the level of a person's retirement savings may depend on the level of a default rate. This idea is captured theoretically by the concepts of "frame" (see Salant and Rubinstein (2008)) or "ancillary condition" (see Bernheim and Rangel (2009)). As suggested in various examples (see e.g. Camerer et al. (2003), and Thaler and Sunstein (2003, 2008)), external conditions can sometimes be chosen by the mechanism designer and, if so, it must be done wisely. Section 8 concludes the paper with a discussion of this issue in relation to my other results.

To summarize, the scope of implementation theory has so far been limited to problems where individual choices can be rationalized via context-independent preferences, while numerous studies have shown this to be unrealistic in various circumstances. Fortunately, this restriction can be dispensed with. Classic necessary and sufficient conditions extend. Though not transparent enough to understand at once which rules are implementable, they can be used effectively to help mechanism designers work around specific biases to implement new non-trivial SCRs. These conditions also played a key role in identifying natural extensions of classic concepts, such as the core or Pareto efficiency, that are Nash implementable on all domains.

Related Literature

The importance of taking behavioral biases into account when designing mechanisms is attracting attention in the popular press, especially since Thaler and Sunstein's (2008) book "Nudge." This interest is also apparent in the academic literature, with an effort to adapt models in industrial organization to determine the best contracts that a monopolist or competing firms can offer to maximize their profits when customers are subject to specific choice biases (see Spiegler's (2011) book for a synthesis).

The present paper extends implementation theory to problems where individual choices may violate IIA, and investigates applications involving specific classes of behavioral biases. These applications are rooted in recent progress made in choice theory to better understand bounded rationality, for instance models of choice with consideration sets (Cherepanov et al. (2009), Manzini and Mariotti (2012), Masatlioglu et al. (2012)), or the model of choice from lists (Rubinstein and Salant (2006)). Proposition 1, which offers necessary and sufficient conditions for Nash implementability, is the only result in this paper that has some precedence in the literature. Indeed, Hurwicz (1986) was the first to highlight that Maskin's (1999) result extends beyond rational choice. Unfortunately, his extension is pertinent only when individual choice functions are multi-valued,[6] while many standard models in choice theory involve single-valued functions, the topic of this paper. Korpela (2012) independently derived a sufficient condition for Nash implementability that is similar to the one I derive in Proposition 1.[7] Independently of both Korpela and I, Ray (2010) derived necessary conditions for Nash implementability. Closer to Hurwicz, his results also boil down to Maskin's standard result when individual choice functions are single-valued, and are thus not useful in my framework.

The papers by Hurwicz, Korpela and Ray extend Maskin's result on Nash implementability in various directions, but do not study what the implications

---

[6]Indeed, his notion of "generalized Nash equilibrium" presumes that each individual picks options that are undominated for an underlying relation that may be intransitive and/or incomplete. Yet, such choice procedures deliver single-valued choice functions only if the underlying relations are complete and transitive, that is only if participants are rational.

[7]I thank Rene Saran for the reference.

may be in relevant applications (such as those described in (i) to (iv) above, for instance). As far as I know, the topic of implementation in dominant strategies has never been discussed in the past in the absence of IIA.

A few other relevant references include Eliaz (2002), who studies full implementation in Nash equilibrium that is robust to the presence of any number of "faulty" individuals below a fixed threshold, where faulty individuals may behave in any possible way; Saran (2011), who studies under which conditions over individual choice correspondences over Savage acts does the revelation principle hold for weak Nash implementation with incomplete information; and Glazer and Rubinstein (2011), who introduce a mechanism design model in which both the content and framing of the mechanism affect the agent's ability to manipulate the information he provides.

2 Definitions

Let I be the set of *individuals*, let Θ be the set of possible *states*, let X be the (finite) set of available *options*, let $P(X)$ be the set of non-empty subsets of X, and let $C_i(\cdot, \theta) : P(X) \to X$ be i's *choice function* when the state is θ, with $C_i(S, \theta) \in S$, for all $S \in P(X)$. Individuals have *private values* if there exist *type sets* $(\Theta_1, \ldots, \Theta_I)$ such that $\Theta = \Theta_1 \times \ldots \times \Theta_I$ and $C_i(\cdot, \theta) = C_i(\cdot, \theta')$, for all θ, θ' such that $\theta_i = \theta'_i$. In that case, i's choice function will be denoted $C_i(\cdot, \theta_i)$. A choice function $C : P(X) \to X$ is *rational* if there exists a strict preference $\succ$ such that $C(S) = \arg \max_\succ S$, for each subset S of X. As is well-known, a choice function C is rational if and only if it satisfies the property of "independence of irrelevant alternatives" (IIA): $C(R) = C(S)$ for all subsets R, S of X such that $R \subseteq S$ and $C(S) \in R$.

A *social choice rule* (SCR) is a correspondence $f : \Theta \to X$ that selects a subset of options for each state. A *mechanism* is a profile $((M_i)_{i \in I}, \mu)$ where M_i is the (finite) set of messages available to i, and $\mu : M \to X$ is the outcome function ($M := \times_{i \in I} M_i$).

A *strategy* for individual i is simply the choice of a message in M_i. In a Nash equilibrium, each player's strategy must deliver his most-preferred

option within the set of options he can generate by varying his own strategy, given that others play their part of the equilibrium. This definition admits a straightforward extension to contexts involving individual choice functions that may violate IIA (in which case one cannot use revealed preferences as Nash did): each individual's strategy leads to an outcome that coincides with his choice within the set of outcomes he can generate by varying his own strategy, while others play their part of the equilibrium. Formally, a profile of strategies $(m_i^*(\theta))_{i \in I}$ forms a *Nash equilibrium* of the game induced by the mechanism (M, μ) at a state θ if

$$\mu(m^*(\theta)) = C_i(\{\mu(m_i, m_{-i}^*(\theta))|m_i \in M_i\}, \theta), \text{ for all } i \in I. \tag{1}$$

When IIA is satisfied and thus revealed preference exists, a strategy is dominant for a player if, whenever combined with arbitrary strategies for the others, it delivers his most-preferred option within the set of options he can generate by varying his own strategy, while others play this arbitrary strategy profile. This definition also admits a straightforward extension to contexts involving individual choice functions that may violate IIA: a strategy is dominant for a player if, whenever combined with arbitrary strategies for the others, it deliver his choice within the set of options he can generate by varying his own strategy, while others play this arbitrary strategy profile. Formally, a message m_i is a *dominant strategy* for individual i at θ if

$$\mu(m_i, m_{-i}) = C_i(\{\mu(m_i', m_{-i})|m_i' \in M_i\}, \theta), \text{ for all } m_{-i} \in M_{-i}.$$

Any profile of dominant strategies thus forms a Nash equilibrium.

The mechanism (M, μ) *implements* the SCR f *in Nash equilibrium* if it admits a Nash equilibrium at every state, and

$$f(\theta) = \{\mu(m^*(\theta))|m^*(\theta) \text{ is a Nash equilibrium at } \theta\}$$

for all $\theta \in \Theta$. If such a mechanism exists, then f is said to be *Nash implementable*. The mechanism (M, μ) *implements* the SCR f *in dominant strategies* if it admits a dominant strategy profile at every state θ, and

$$f(\theta) = \{\mu(m^*(\theta))|m^*(\theta) \text{ is a dominant strategy profile at } \theta\},$$

9

for all $\theta \in \Theta$. If such a mechanism exists, then f is said to be *dominant strategy implementable.*

3 Necessary and Sufficient Conditions

This section offers abstract necessary and sufficient conditions for implementability in either sense. They will prove useful in subsequent sections to determine which SCRs are implementable in relevant applications.

3.1 Nash Implementation

I restrict attention to single-valued SCRs. Some more general results are provided in the supplemental appendix.

Let's start by reminding the essence of Maskin's (1999) result. "Monotonicity" is the key property that emerged from his work.

Maskin Monotonicity *Suppose that $C_i(\cdot, \theta)$ is rational, for each $i \in I$ and each $\theta \in \Theta$. Then a single-valued SCR $f : \Theta \to X$ is Maskin Monotonic if $f(\theta') = f(\theta)$, for each θ, θ' such that $LC_i(f(\theta), \theta) \subseteq LC_i(f(\theta), \theta')$ for each $i \in I$, where $LC_i(x, \theta) = \{y \in X | x \succ_i (\theta) y\}$ denotes the lower contour set of x given the preference ordering $\succ_i (\theta)$ associated to $C_i(\cdot, \theta)$.*

Though necessary, the property is not quite sufficient, and various conditions have been proposed over the years to complete it so as to guarantee Nash implementability. Perhaps the simplest and most well-known remains Maskin's "No Veto."

No Veto *Suppose that $C_i(\cdot, \theta)$ is rational, for each $i \in I$ and each $\theta \in \Theta$. Then a single-valued SCR $f : \Theta \to X$ satisfies no veto if $f(\theta) = x$, for each $\theta \in \Theta$ for which there exists $i \in I$ such that x is top-ranked according to $\succ_j (\theta)$, for each $j \in I \setminus \{i\}$.*

Maskin's (1999) classic result establishes that the two following results hold on any domain that contain only rational choice functions: 1) a SCR is Nash implementable only if it is Makin Monotonic, and 2) any SCR that is both Maskin Monotonic and satisfies No Veto is Nash implementable if $\#I \geq 3$.

The following property offers one straightforward extension of No Veto to more general individual choice functions.

Property N *Let $\theta \in \Theta$. If there exist x and $i \in I$ such that $C_j(X, \theta) = x$, for all $j \in I \setminus \{i\}$, then $f(\theta) = x$.*

The question of how to extend the more substantial property of Maskin Monotonicity is of greater interest. I show in the supplemental appendix that two natural definitions, a monotonicity property based on Bernheim and Rangel's (2009) extended revealed preference and the choice-based monotonicity property proposed by Aizerman and Aleskerov (1986) in their study of the aggregation of individual choice functions into social choice functions, are both inappropriate for characterizing implementable SCRs.

Here is another property that coincides with Maskin Monotonicity when individual choice functions satisfy IIA, and that will be useful for my purpose in that it is necessary for Nash-implementability, and sufficient once combined with Property N.

Property M *For each i, there exists a function $X_i : \Theta \to P(X)$ such that*

1. *$C_i(X_i(\theta), \theta) = f(\theta)$, for all $\theta \in \Theta$ and all $i \in I$,*

2. *For all θ, θ', if $C_i(X_i(\theta), \theta') = f(\theta)$, for all i, then $f(\theta') = f(\theta)$.*

The property requires that, for each state θ and each individual i, one can find a set $X_i(\theta)$ of alternatives such that i picks $f(\theta)$ out of $X_i(\theta)$ when the state is θ. In addition, if all individuals pick $f(\theta')$ out of $X_i(\theta)$ when the state is θ', then $f(\theta')$ must coincide with $f(\theta)$. Consider now the special case where individual choice functions satisfy IIA at all states. If Property M is satisfied for some profile $(X_i)_{i \in I}$, then so is it for any other profile $(X_i')_{i \in I}$ that is larger (i.e. $X_i(\theta) \subseteq X_i'(\theta)$, for each i and θ), provided that the first condition in Property M remains satisfied. Hence Property M is equivalent to that same property applied to the largest profile for which the first condition remains valid, i.e. $X_i(\theta) = LC_i(f(\theta), \theta)$, and the second condition in Property M indeed boils down to Maskin monotonicity in that case.

One is now ready to extend Maskin's result to any domain: 1) a single-valued SCR is Nash implementable only if it satisfies Property M, and 2)

any single-valued SCR that satisfies both M and N is Nash implementable if $\#I \geq 3$. Yet, it will be useful in applications below to have slightly weaker sufficient condition. Multiple papers have been written to characterize Nash-implementability on rational domains. My objective here is not to find extensions of these various results to my more general framework. Instead, I propose a condition inspired from these papers that is adequate to cover all the applications discussed in the rest of the paper. Finding a single condition that would be both necessary and sufficient for Nash-implementability on any domain remains an open question of theoretical interest.

Property $\mathfrak{M}$ *There exists a function $X_i : \Theta \rightarrow P(X)$, for each $i \in I$, such that the two conditions in Property M are satisfied, and if there exists $j \in I$, $\theta, \theta' \in \Theta$, and $x \in X$ such that $C_i(X, \theta) = x$, for all $i \in I \setminus \{j\}$, and either $C_j(X, \theta) = x$ or $C_j(X_j(\theta'), \theta) = x$, then $f(\theta) = x$.*

The restriction added to M to get $\mathfrak{M}$ are weaker than N, since it applies only under some additional condition on j's choice.

Proposition 1 (NECESSARY, AND SUFFICIENT CONDITIONS FOR NASH-IMPLEMENTABILITY) *Let f be a single-valued SCR. If f is Nash implementable, then it satisfies Property M. If $\#I \geq 3$, and f satisfies Property $\mathfrak{M}$, then it is Nash implementable.*

3.2   Dominant Strategy Implementation

I restrict attention to private values in this section. I start with an extension of the revelation principle to my more general domain where states encode individual choice functions, in order to characterize the set of SCRs that are implementable in dominant strategies.

Proposition 2 (WEAK REVELATION PRINCIPLE) *Consider a problem with private values. The SCR f is implementable in dominant strategies if and only if it is single-valued, and there exists a set $\mathcal{M}_i$ of messages, for each $i \in I$, such that truth-telling is a dominant strategy for every type profile in a mechanism where individual i's message space is $\Theta_i \cup \mathcal{M}_i$, for each $i \in I$, and the outcome function coincides with f when all individuals report types.*

This weak revelation principle highlights that, as is the case when individual choice functions satisfy IIA, there is no loss of generality for the mechanism designer to make his decision on which option to implement based solely on the individuals' direct reports about their types. On the other hand, the fact that individuals' choices may violate IIA implies that the mechanism designer may want to add some 'decoy messages' to 'nudge' individuals to see truth-telling as a dominant strategy. Obviously, Proposition 2 boils down to the classic revelation principle when individual choice functions satisfy IIA, as messages in $\mathcal{M}_i$ can be deleted without changing the individuals' behavior in the mechanism. Note that the exact same class of mechanisms was introduced by Mookherjee and Reichelstein (1990) under the name of "augmented revelation mechanism" to state their "augmented revelation principle." They show that any SCR which is Nash implementable (on rational domains) can be implemented by an augmented revelation mechanism for which truthtelling is one equilibrium, and then state the "selective elimination condition" to guarantee that truth is the only equilibrium, thereby providing a sufficient condition for Nash implementability. The non-type auxiliary messages are used there to destroy undesired equilibria without introducing new ones, while here they are used to influence individual choice behaviors.

The fact that $\mathcal{M}_i$ may have to be non-empty makes this weak revelation principle less helpful than its more standard version in the special case where preferences are rational. Phrasing a tractable condition that is both necessary and sufficient seems out of reach. Yet the weak revelation principle allows to derive simpler necessary and sufficient conditions that share some similarities with Property M, and will prove useful in applications below.

Property P *For each $i \in I$, there exists a function $X_i : \Theta_{-i} \to P(X)$ such that $C_i(X_i(\theta_{-i}), \theta_i) = f(\theta)$, for all $\theta \in \Theta$.*

The property requires that, for each i and each profile θ_{-i} of types for the other individuals, one can find a set $X_i(\theta_{-i})$ of alternatives such that i picks $f(\theta)$ out of $X_i(\theta_{-i})$ when the state is θ.

Proposition 3 (Necessary, and Sufficient Conditions for Imple-

MENTABILITY IN DOMINANT STRATEGIES*) *Consider a problem with private values. If a SCR f is implementable in dominant strategies, then it satisfies Property P. Conversely, if a SCR f satisfies property P for the specific sequence of functions $(X_i^*)_{i \in I}$, where $X_i^*(\theta_{-i}) = \{f(\theta_i, \theta_{-i}) | \theta_i \in \Theta_i\}$, for each θ_{-i} and each i, then f is implementable in dominant strategies.*

Despite their slight resemblance, there is no systematic logical relation between M and P (and no general logical implication between Nash-implementation and implementation in dominant strategies). The fact that M does not always implies P is already true on some domains containing only rational choice functions (see Theorem 7.2.3 and the discussion that follows it in Dasgupta et al. (1979)). P implies M on domains that contain only individual choice functions that satisfy IIA (see Theorem 3.2.1 in Dasgupta et al. (1979)), but I provide an example in the supplemental appendix to show that this logical implication breaks down when considering more general individual choice functions.

4   Efficiency

The Pareto correspondence is one of the most standard examples of SCR that is Nash implementable[8] on rational domains, but unfortunately is not well-defined without the existence of a revealed preference (i.e. in the absence of IIA). One possible extension of the Pareto correspondence is proposed by Bernheim and Rangel (2009). Following their definition, an option a is *unambiguously preferred* to an alternative b given a choice function C if $C(S) \neq b$, for all S such that $a \in S$. Though necessarily incomplete when IIA is violated, this revealed ordering and the associated Pareto principle may still allow to compare some options. An option is *BR-efficient* if there is no alternative option that is unambiguously preferred by all individuals. Let F^{BRE} be the SCR that associates to each state θ the set of BR-efficient options at θ.

I now suggest an alternative extension of Pareto efficiency in the absence of IIA. To motivate the definition, notice that the Pareto criterion for rational

---

[8]The paper restricts attention to single-valued individual choice functions (equivalent to focusing on strict preferences on rational domains). It is impossible in that case to implement multi-valued SCRs in dominant strategies (as observed in Proposition 2).

preferences can be rephrased in terms of "freedom of choice" instead of revealed preferences. Picking an option x could be interpreted as implicitly giving each individual i the freedom to choose from a subset that contains x and is contained in its lower contour set. As Pareto efficiency has no implications in terms of fairness, these lower contour sets (one for each individual) can be relatively small or relatively large. Yet it cannot be that all of them are small. More precisely, an option x is Pareto efficient if and only if one can find an implicit choice set for each individual out of which he would pick x, and such that each alternative belongs to the implicit choice set of at least one individual. Indeed, y is a Pareto improvement over x if and only if one cannot find an individual whose lower contour set for x contains y. The argument in terms of freedom of choice instead of revealed preference has the advantage of admitting a straightforward extension for individual choice functions that need not satisfy IIA. This leads to an extension of the Pareto correspondence that is different from F^{BRE}. An element x is efficient according to F^{Eff} if one can associate to each individual i an implicit choice set Y_i such that, for each alternative $y \in X \setminus \{x\}$, there is at least one individual who was free to choose y, but decided to pick x instead. Formally.

$$F^{Eff}(\theta) = \{x \in X | (\exists (Y_i)_{i \in I} \text{ subsets of } X) : C_i(Y_i, \theta) = x, \text{ for all } i, \text{ and } X = \cup_{i \in I} Y_i\},$$

for each $\theta \in \Theta$. Notice that F^{Eff} has non-empty values, as needed, since it includes any option x such that $x = C_i(X, \theta)$, for some $i \in I$.

Proposition 4 *Suppose that X contains at least three elements. F^{Eff} is Nash implementable on all domains, while F^{BRE} isn't. $F^{Eff} \subseteq F^{BRE}$, and F^{Eff} coincides with the set of Pareto optimal options on rational domains.*

An advantage of the definition of F^{Eff} (and F^{BRE}) is that it is phrased exclusively in terms of observables (i.e. choices), and is thus independent of a specific model or way to interpret choices. Of course, if a modeler strongly believes in a model (which requires at the very least that this model is not contradicted by observed choices), then he may well disagree with the fact that F^{Eff} captures a reasonable notion of efficiency. For instance, I will discuss in Section 7 models where individuals make choices by maximizing a well-defined

preferences, as rational agents would do, but need not pay attention to all feasible alternatives. If one believes that this is the way choices are determined, then one may reject the idea that observing an individual i picking x out of Y_i means that i had the freedom to choose any alternative option y in Y_i. Arguments along these lines can be found in Salant and Rubinstein (2008) and Masatlioglu et al. (2012). I study in the supplemental appendix the correspondence that associates to each state the set of Pareto efficient allocations for the true underlying preferences in the model of choice with consideration sets. Unfortunately I show that it is not always Nash implementable either. Independently of whether one finds F^{Eff} a desirable SCR or not, it remains a fact that it is Nash implementable on all domains.

As F^{Eff} constitutes a selection of Bernheim and Rangel's (2009) extended notion of efficiency, it is important to understand why options that are considered efficient according to their definition need not be according to mine. Consider a case where available options are labeled according to the letters in the alphabet. All individuals except the first have a rational choice function that coincides with the natural ordering. The first individual has the same choice function except that he picks option z out of any pair that contains it. Option z is BR-efficient, since z is not unambiguously dominated by any alternative for the first individual. Yet it is not efficient according to my definition. Indeed, the implicit choice set for all individuals different from the first must be the singleton $\{z\}$, as this is the worst option for them. The implicit choice set for the first individual must be a pair that contains z given his choice function. Hence, for any profile of implicit choice sets $(Y_i)_{i \in I}$, most feasible alternatives do not belong to any implicit choice set if each individual i picks z out of Y_i.

5   Core in House Allocation

The core is another important example of SCR that is Nash implementable when individual choice functions satisfy IIA. In this section, I present results on this topic without requiring IIA in the context of Shapley and Scarf's (1974)

house allocation problem, a classic matching environment.

Each individual owns one unit of an indivisible object. Let O denote the set of all objects: $O = \{o_i^* | i \in I\}$, where o_i^* is i's initial endowment. Then $X = \{z \in O^I | z_i \neq z_j, \forall i \neq j\}$. To avoid individual choice correspondences, while assuming that individuals care only about the object they consume, choices are defined in this section over subsets of O instead of X. So each state θ defines a choice function $C_i(\cdot, \theta) : P(O) \to O$ for each individual i. As before, a SCR is a function that associates an element of X to each element of Θ, but elements in the range of f are now vectors with I components, given the specific structure of X. A similar point applies to the outcome function of mechanisms. Condition (1) defining a Nash equilibrium thus now becomes:

$$\mu_i(m^*(\theta)) = C_i(\{\mu_i(m_i, m_{-i}^*(\theta)) | m_i \in M_i\}, \theta).$$

For each group S of individuals, let then $\mathcal{F}(S)$ be the set of house allocations that can be achieved by its members. Formally, $\mathcal{F}(S)$ is the set of $\alpha \in O^S$ such that $\{\alpha_i | i \in S\} = \{o_i^* | i \in S\}$ and $\alpha_i \neq \alpha_j$ whenever $i \neq j$. Inspired by the definition of F^{Eff}, consider the following SCR:

$$F^C(\theta) = \{x \in \mathcal{F}(I) | (\exists (T_i)_{i \in I} \text{ subsets of } O) : C_i(T_i, \theta) = x_i, \forall i, \text{ and}$$

$$(\forall S \in P(I))(\forall \alpha \in \mathcal{F}(S))(\exists i \in S) : \alpha_i \in T_i\}$$

In order to make sure that F^C is a well-defined SCR, I need to check that it is non-empty valued. For this, consider the following extension of Shapley and Scarf's (1974) *top trading cycle* procedure. Fix a state θ, and compute at that state the object that each individual i would pick if he was free to choose any element of O. Draw an arc from individual i and j if and only if $C_i(O, \theta)$ is j's endowment. This graph must have at least one cycle. Implement the trades induced by all these cycles, and let $O' \subsetneq O$ be the set of objects remaining, i.e. $O' = \{o_i^* | i \in S\}$ where S is the set of individuals who weren't part of a cycle. Then iterate the procedure with the remaining individuals and objects. Call $\alpha^*(\theta)$ the resulting allocation of objects. Let's check that $\alpha^*(\theta) \in F^C(\theta)$, for each $\theta \in \Theta$. For each individual i, let T_i be the set of objects remaining when he became part of a trading cycle in the procedure.

The fact that $C_i(T_i, \theta_i) = \alpha_i^*(\theta)$ thus follows immediately from the definition of that cycle in the procedure. Let $S \in P(I)$, $\alpha \in \mathcal{F}(S)$, and let j be one of the individuals in S who is among the earliest to be part of a trading cycle in the induction. Then $\alpha_j \in T_j$, for each $\alpha \in \mathcal{F}(S)$, by definition of the induction, of $\mathcal{F}(S)$ and of T_j. I have thus proved indeed that $\alpha^*(\theta) \in F^C(\theta)$.

Bernheim and Rangel's (2009) extended revealed preference can also be used to define an extension of the core that is applicable in particular to the house allocation problem. Say that coalition S *BR-blocks* an allocation α if there exists $\beta \in \mathcal{F}(S)$ such that β_i unambiguously dominates α_i, for each $i \in S$. The *BR-core*, denoted F^{BRC}, then associates to each profile θ of types the set of allocations $\alpha \in \mathcal{F}(I)$ such that there is no coalition S that BR-blocks α. I now provide a result in the spirit of Proposition 4.[9]

Proposition 5 *Suppose that O contains at least three elements. F^C is Nash implementable on all domains, while F^{BRC} isn't. $F^C \subseteq F^{BRC}$, and F^C coincides with the core on rational domains.*

Remark 1 *If individual choice functions satisfy IIA, then F^C coincides with the core, as just observed, and hence the allocation resulting from the top trading cycle is the only element in the core (which also coincides with the unique competitive equilibrium). However, $\alpha^*(\theta)$ need not be the only element of $F^C(\theta)$ for states associated to individual choice functions that may violate IIA. An example in that regard is provided in the supplemental appendix.*

6   Rich Domains

An option x is said to be *most preferred* according to a choice function C if $C(S) = x$, for every S that contains x. Given a most preferred option x, an

---

[9]Bade (2008) studies the house allocation problem in the presence of individual choice functions that need not satisfy IIA. She defines two notions of efficiency and of cores, one based on Bernheim and Rangel's (2009) revealed preference, cf. F^{BRE} and F^{BRC}, and one based on an alternative revealed ordering whereby $a \succ b$ if there exists S that contains b and such that $C(S) = a$. The latter ordering is complete but often intransitive, leading in many cases to the emptiness of both concepts (efficiency and core). She does not study the implementability of the various concepts she defines, but instead defines the top-trading cycle (same as above) and serial dictatorship procedures to show that an equivalence result by Abdulkadiroglu and Sonmez (1998) does not extend to the case where individual choice functions violate IIA.

option y is said to be *second most preferred* according to a choice function C if $C(S) = y$, for all S that contains y but not x. Given a most preferred option x, and a second most preferred option y, an option z is said to be *third most preferred* according to a choice function C if $C(S) = z$, for all S that contains z but not x, nor y. Rational choice functions systematically admit a most, second most, and third most preferred option, but violations of IIA does not necessarily preclude the existence of such options either.

In this section, I restrict attention to problems with private values. A single-valued SCR $f : \Theta \to X$ has a *rich domain* if for each $x, y, z \in Range(f)$ and each individual i, there is a type $\theta_i \in \Theta_i$ such that x is most preferred according to $C_i(\cdot, \theta_i)$, y is second most preferred, and z is third most preferred. Having a rich domain thus means that the mechanism designer cannot exclude[10] that participants have a most, a second most, and a third most preferred option, and that any three options in the range of f can play these roles (no restriction in "tastes"). This boils down to a simplified version of Kalai et al.'s (1979) free triple condition when individual choice functions satisfy IIA. Classic impossibility results in such case imply that dictatorial rules are the only Pareto efficient SCRs with a range of at least three elements that are either Nash implementable or implementable in dominant strategies. Here is an extension of these results to numerous new domains that include types associated to individual choice functions that violate IIA.

Proposition 6 *Consider a problem with private values. Let $f : \Theta \to X$ be a single-valued SCR that has a rich domain, is BR-efficient,[11] and whose range contains at least three elements. If f is either Nash implementable or implementable in dominant strategies, then there exists a unique j and $X_j :$ $\Theta \to P(X)$ such that $f(\theta) = C_j(X_j(\theta), \theta_j)$ and $Range(f) \subseteq X_j(\theta)$, for each θ.*

Individual j in the previous proposition qualifies as a dictator insofar as the corresponding SCR picks x whenever x is his most-preferred option. The simplest example of SCR meeting the assumptions of Proposition 6 is given

---

[10]"Cannot exclude" means that these different properties are met by some types.

[11]The result remains true if BR-efficiency is replaced by the alternative notion I introduced in Section 4, since $F^{Eff} \subseteq F^{BRE}$.

by $f(\theta) = C_j(Range(f), \theta_j)$. Yet, violations of IIA make it possible for more subtle SCRs to be implementable as well. Indeed, notice $X_j(\theta)$ can possibly vary with θ_{-i}. Here is indeed an example. Consider the hiring problem with attraction effect discussed in the Introduction. Let Z be a set of candidates who are above the bar, out of which members of a committee I must hire one applicant. For every $z \in Z$, there exists a candidate $\gamma(z) \in X \setminus Z$ that is similar to z but clearly ranked below the bar. Having the option to pick $\gamma(z)$ may nudge some participants to pick z out of $Z \cup \{\gamma(z)\}$, in the spirit of the attraction effect (Huber et al. (1982)). Those who are not subject to that effect are assumed to overlook the feasibility of $\gamma(z)$. Formally, the type sets are assumed to satisfy the following conditions. For each $i \in I$, there exists a subset A_i of types that are subject to the attraction effect:

1. if $\theta_i \in A_i$ and $z \in Z$, then $C_i(Z \cup \{\gamma(z)\}, \theta_i) = z$,
2. if $\theta_i \in \Theta_i \setminus A_i$ and $z \in Z$, then $C_i(Z \cup \{\gamma(z)\}, \theta_i) = C_i(Z, \theta_i)$.

Given that $\gamma(z)$ is similar to z, it will also be required that $C_i(\gamma(Z), \theta_i) = \gamma(C_i(Z, \theta_i))$, for all $\theta_i \in \Theta_i$ and all $i \in I$. Though not needed, I will also assume that there exists an option $c \in X$ that is so bad or dissimilar that it does not influence choices out of subsets of Z independently of individuals' types: $C_i(S \cup \{c\}, \theta_i) = C_i(S, \theta_i)$, for all subset S of Z and all θ_i. This will greatly simplify arguments in the proof of Nash implementability in the next proposition. It is easy to see that numerous domains that are rich over Z satisfy these assumptions.

For each $i, j \in I$, consider the following SCR: $f_{ij}(\theta) = C_i(Z, \theta_i)$ if $\theta_i \in \Theta_i \setminus A_i$, and $f_{ij}(\theta) = C_j(Z, \theta_j)$ if $\theta_i \in A_i$. In other words, individual i determines the outcome when he is not subject to the attraction effect. Otherwise individual j does. The range of f_{ij} is Z whenever the domain is rich over Z. Also, $f_{ij}(\theta) \in F^{Eff}(\theta) \subseteq F^{BRE}(\theta)$, for all θ. Notice how individual j's type plays a role in determining the final outcome being implemented, despite the fact that i qualifies as a dictator according to my earlier definition. The SCR f_{ij} is implementable both in Nash equilibrium and in dominant strategies.

Proposition 7 *Let i, j be any two individuals. Then the SCR f_{ij} is implementable both in Nash equilibrium and in dominant strategies.*

7 Applications to Specific Classes of Biases

This section studies implementation on domains where individual choices can be described by specific models capturing structured departures from rationality. Results in Sections 4 and 5 establish that some important multi-valued SCRs are Nash implementable. Proposition 6, on the other hand, establishes that few single-valued SCRs are implementable (either in Nash equilibrium or in dominant strategies) when the domain is rich. In perspective of these results, the question is whether there are ways to exploit specific biases to define non trivial single-valued SCRs that are implementable. I focus on three applications: problems with fixed consideration sets, problems with lists, and problems involving individuals who fight temptation with limited willpower. The associated domains of choice functions need not be rich, even when allowing for a diverse range of behavior within those classes of biases. While results remain strongly negative on the first domain, they are more permissive on the other two.

7.1 Fixed Consideration Sets

In this subsection, I focus on problems where individuals are maximizing a preference relation, but may overlook some feasible alternatives when picking their choices. Such an idea was first introduced in the marketing literature, where the feasible set describes the set of all options the individual is aware of, while the consideration set is the subset of those options that the individual actively considers when making his decision (see e.g. Wright and Barbour (1977) and Roberts and Lattin (1991), see also Lleras et al. (2010) and Masatlioglu et al. (2012) for a recent elaboration on this idea in choice theory). Similarly, in Cherepanov et al. (2009), each individual is assumed to maximize his preference over the subset of options that he can "rationalize," in the sense of being maximal within the feasible set for at least one "rationale." Manzini and Mariotti (2012) offer yet another motivation for the framework studied in this subsection. They suggest that people may overlook dominated categories of options when making choices out of large sets. For instance, when having to choose among many restaurants, one may first decide to focus only on some

specific cuisine (e.g. italian), and then pick the best option within that category, while one would consider all the options (and perhaps prefer a mexican restaurant over an italian alternative) when there are only a handful of them. Similarly, one may restrict attention to apartments in a one mile-radius from one's workplace when there are numerous apartments available for rent, but would drop that restriction if only few are available. The model of choice from lists introduced in the next subsection provides additional examples, if individuals pay attention to the first $k(s)$ elements of any list with s elements. The key assumption underlying the results in this subsection is that consideration sets do not vary with the state.

Restricting attention to problems with private values, I assume in addition that, for each individual i, there exists a *consideration correspondence* $A_i : P(X) \to P(X)$, associating to each set $S \subseteq X$ of feasible options the subset $A_i(S) \subseteq S$ of options that he actually considers, such that

$$\forall \theta_i \in \Theta_i, \exists \theta_i^* \in \Theta_i^*, \forall S \in P(X) : C_i(S, \theta_i) = \arg\max_{\succ(\theta_i^*)} A_i(S), \qquad (2)$$

where Θ_i^* denotes the set of rational types. In other words, individual i picks the option that maximizes $\succ(\theta_i^*)$ over $A_i(S)$. While consideration sets are fixed in this subsection, underlying preferences/tastes are left unrestricted:

$$\forall \theta_i^* \in \Theta_i^*, \exists \theta_i \in \Theta_i, \forall S \in P(X) : C_i(S, \theta_i) = \arg\max_{\succ(\theta_i^*)} A_i(S). \qquad (3)$$

Despite the fact that preferences are unrestricted, domains derived from conditions (2) and (3) are usually not rich, and so Proposition 6 does not apply. Indeed, an option that is feasible but not considered in some choice problem is not most preferred according to the definition of the previous section. Does implementation become more permissive in societies with individuals who have different consideration sets? Unfortunately the answer is negative.

Proposition 8 *Consider a problem with private values which satisfies conditions (2) and (3). Let f be a single valued SCR whose range contains at least three options. Then f is Nash implementable (implementable in dominant strategies, respectively) if and only if there exists $j \in I$ and $S \in P(X)$ such that $Range(f) = A_j(S)$ and $f(\theta) = C_j(S, \theta_j)$, for all $\theta \in \Theta$.*

Consideration sets do not vary with types in this subsection. The negative

conclusions derived in the previous proposition need not apply otherwise. For instance, the hiring problem with an attraction effect from the previous section could be rephrased with types specifying preferences and consideration sets, and yet Proposition 6 shows that there exist implementable SCRs that vary with the type of multiple individuals. The next two subsections contain other instances of more permissive results because consideration sets vary with states.

7.2   Choice from Lists

I now present a more permissive implementation result for specific classes of choice procedures defined on lists. Assume, for instance, that the mechanism designer has labeled the various options as "Option 1 is $(\dots)$", "Option 2 is $(\dots)$", etc. This establishes a natural *list l*, which simply amounts to an enumeration of the elements of X, $(x_k)_{k=1}^{|X|}$. It is then assumed that, for any feasible set, individuals investigate the relative appeal of its various elements in an order that is consistent with the mechanism designer's labeling system. A first example of choice procedure I will consider is given by Rubinstein and Salant's (2006, Example 6) "Stop When You Start to Decline." In this model, individuals are endowed with a complete and transitive strict preference ordering on X (which, in the present paper, may vary with their types), and check options out of any feasible set $S \subseteq X$ by following the sequence that is consistent with the list's enumeration. They continue this process as long as the next option improves upon the last according to their preference. The option picked from S is then the last in this improving sequence, i.e. the last one they paid attention to before finding a new option that is worse according to their preference. A second example of choice procedure covered in this subsection, is Simon's (1955) satisficing procedure, as formalized in Rubinstein and Salant (2006, Example 2). In this alternative model, individuals are endowed with a utility function and a real number that represents a threshold (both of which may now vary in with their types). Again, they are assumed to check options out of any feasible set $S \subseteq X$ by following the sequence that is consistent with the list's enumeration. They then pick the first option whose utility

is larger than the threshold. If no option meets this requirement, then they pick the last option they have looked at among those that are feasible. The satisficing procedure leads to choices that happen to satisfy IIA, but whose revealed preferences belong to a class that has not been studied so far in the implementation literature. "Stop When You Start to Decline" leads to choices that violate IIA, and thus are not compatible with preference maximization.

More generally, I will consider in this subsection any domain Θ with private values such that type profiles satisfy the following condition: for each i and each $\theta_i \in \Theta_i$, if there exists $K \in \{1, \ldots, |X|\}$ such that $S = \{x_1, \ldots, x_K\}$, then

$$
C_i(S, \theta_i) = \begin{cases} C_i(X, \theta_i) & \text{if } C_i(X, \theta_i) \in S \\ x_K & \text{if } C_i(X, \theta_i) \notin S. \end{cases} \tag{4}
$$

This condition only restricts the way people choose out of subsets of the form $\{x_1, \ldots, x_K\}$ for some $K \in \{1, \ldots, |X|\}$. It is thus a weaker requirement than IIA. On the other hand, the condition $C_i(S, \theta_i) = x_K$ if $C_i(X, \theta_i) \notin S$ imposes a restriction on choices that isn't covered by IIA. It is easy to check that the two choice procedures presented in the previous paragraph lead to individual choice functions that satisfy condition (4).

Remark 2 *If individuals are rational, and have single-peaked preferences on the list, then they satisfy condition (4). Notice also that choice functions generated by applying the satisficing procedure satisfy IIA, and can thus be rationalized by the maximization of a complete transitive strict ordering, but that this underlying preference need not be single-peaked.*

Let f^l be the (single-valued) SCR that selects the first option in the list that is picked out of X by at least one individual: $f^l(\theta) = x_{k^l(\theta)}$, where $k^l(\theta) = \min\{k \in \{1, \ldots, |X|\} | (\exists i \in I) : x_k = C_i(X, \theta_i)\}$. It is easy to check that it is anonymous, has full range, selects elements of F^{Eff}, and thus a fortiori guarantees BR-efficiency as well.

Proposition 9 *Consider a domain Θ that satisfies condition (4). Then f^l is Nash implementable if $|I| \geq 3$. It is also implementable in dominant strategies.*

I now establish a weak converse. For each $i \in I$, let $\bar{\Theta}_i$ be the domain of *all* choice functions that satisfy condition (4). Of particular interest is the case where choice functions arise from the 'stop when you start to decline' procedure. Let Θ_D be the subsets of choice functions that can be generated by this procedure (varying the underlying preference to which it is applied). I will focus on (the more restrictive concept of) implementation in dominant strategies via simple mechanisms. Simple mechanisms allow each individual to send a message which is an option in X, i.e. $M_i = X$. Otherwise, the concept of implementation is the same as in Section 2.

Proposition 10 *Let f be a single-valued SCR defined over a domain $\Theta \subseteq \bar{\Theta}$ that includes Θ_D. If f is BR-efficient,[12] anonymous, and implementable in dominant strategies via a simple mechanism, then $f = f^l$.*

7.3   Building Willpower in Groups

Temptation is often understood in economics through the lens of commitment preferences (see Lipman and Pesendorfer (2011) for a survey): an individual who anticipates having to fight temptation at the time of making a choice, may want to commit to a smaller set of options. By contrast, the focus in the psychology literature is often on individuals who make choices while being tempted, and how they deplete their willpower when exercising self-control (see e.g. Baumeister and Tierney (2011)).[13]

Here is a stylized model of choice that captures temptation, self-control and willpower in a way that is closer to the psychology literature. There are n individuals with a common long-term goal. This long-term goal is difficult to achieve due to the presence of tempting alternatives: choices are also influenced by a short-term craving. Each individual has some limited willpower to exercise

---

[12]The result also holds with the weaker requirement that f is Pareto efficient at each type profile for which all individuals are rational.

[13]First attempts to capture willpower in economics include Ozdenoren et al. (2010) who study the optimal management of willpower over time, and Masatlioglu et al. (2011) who study commitment preferences for individuals who anticipate having to fight temptation with limited willpower when having to make a choice.

self-control. In this example, willpower is captured by the number of tempting options an individual can overlook to better fulfill his long-term goal. Formally, if T is the set of feasible options, $\succ_L$ (an ordering of X) captures the long-term goal, $\succ_S$ (another ordering of X) captures the short-term craving, and i's willpower is captured by a positive integer k_i, then i's choice out of T is the most-preferred element for $\succ_L$ among those that are dominated by at most k_i alternatives according to $\succ_S$. Such behavior typically leads to violations of IIA. For instance, one may be able to resist eating a slice of pizza for lunch when the alternative is a salad, but unable to resist the temptation of both the burger and the pizza slice, and go for the slice if these two options are available on the menu in addition to the salad. This choice pattern can be explained if $k_i = 1$, the long-term goal ranks the salad above the pizza slice, and the pizza slice above the salad, while the short-term preference is exactly opposed.

Consider now a situation where a state determines a common long-term goal, and the individuals' (possibly different) short-term cravings. Is there a way to combine the individuals' limited willpower to help them better fulfill their common long-term goal? One idea is to decentralize the burden of choice by letting each individual be 'in charge of' only a small number of alternatives. Here is a simple mechanism to achieve this. For each i, let $A_i \subset X$ be the subset of k_i elements that individual i will be in charge of. Let's assume that $\sum_{i \in I} k_i \geq \#X$. In that case, we can define these sets in such a way that $\cup_{i \in I} A_i = X$. Each individual picks a message in support of one of the elements he is in charge of, as well as a non-negative integer representing the intensity with which he makes that statement. Formally, $M_i = A_i \times \mathbb{Z}_+$. The outcome is the option supported by the individual with most intense report (using any fixed tie-breaking rule if multiple messages are reported with the highest intensity).

The mechanism has the property that, for any message profile, individual i can generate at most $k_i + 1$ options (the outcome associated to that message profile plus those in A_i) when varying his own message, out of which he picks the option that is best for his long-term goal. Clearly, the option

that is top ranked in X for the long-term goal is an equilibrium outcome of this mechanism, e.g. with the individual in charge of it suggesting it with intensity 1 and all other individuals supporting other options with intensity 0. Conversely, any Nash equilibrium outcome must be the best element in X for the long-term goal. Otherwise, the individual in charge of it will deviate by supporting it in a message whose intensity is larger than the messages from all other individuals. We have thus proved the following result.[14]

Proposition 11 *If $\sum_{i \in I} k_i \geq \#X$, then the SCR that systematically selects the top-choice of the common long-term goal is Nash implementable.*

8 Concluding Remark: Frames

Choices can sometimes be influenced by external conditions. For instance, the meal picked in a cafeteria may vary with the order in which options are displayed, or the level of a person's retirement savings may depend on the level of a default rate (see Thaler and Sunstein (2008)). This idea is captured theoretically by the concepts of "frames" (see Salant and Rubinstein (2008)) or "ancillary conditions" (see Bernheim and Rangel (2009)).

The model discussed so far can be amended to encompass this added flexibility. Let Φ be a finite set of frames that the mechanism designer can choose from (e.g. the default saving rate for retirement, the list in which to present items, an initial allocation of houses in the house allocation problem, etc.). Choice functions are now indexed by frames: $C_i^\phi(\cdot, \theta) : P(X) \to X$, for each i and each θ, with $C_i^\phi(S, \theta) \in S$, for all $S \in P(X)$.[15] A SCR $f : \Theta \to X$

---

[14]The proposition can also be proved by defining adequate functions X_i, and checking Property $\mathfrak{M}$. The alternative argument just followed show that the SCR is also implementable via a mechanism that is simpler and more intuitive than the canonical mechanism defined in the proof of Proposition 1. The mechanisms derived either way does not admit a dominant strategy at any state, and hence does not implement the SCR in dominant strategies. I conjecture that it is not implementable in dominant strategies, but this remains an open question.

[15] Bounded rationality now acquires an additional dimension. For instance, an individual's choices may satisfy IIA for each given frame, but be boundedly rational because these choices vary with frames. Many examples of endowment effects, or Rubinstein and Salant's (2006) version of Simon's (1955) satisficing procedure, fall in this category.

is implementable if there exists a frame $\phi \in \Phi$ such that f is implementable in the problem (Θ, C^ϕ). It should be clear that all the results derived so far extend to this problem after the frame has been fixed. Adding frames thus expands the set of SCRs that the mechanism designer can implement, but does not require additional theoretical work.

Since individuals must know the rules of the mechanism before playing it, frames cannot vary with the state. It is well-known that considering dynamic games expands the set of single-valued SCRs that are implementable (in subgame-perfect equilibrium). A new feature when considering dynamic games in the present model is the added possibility of having frames vary endogenously with underlying states. The supplemental appendix contains a characterization of SCRs that are implementable by backward induction in a two-stage mechanism where a fully[16] rational individual moves first, before another individual makes a choice that determines the final outcome. It then offers an example to illustrate how implementation in this sense allows for frames to vary endogenously with states.

---

[16]In the sense of satisfying IIA and being unaffected by frames (e.g. a profit maximizing firm or the head cook of a school cafeteria).

Appendix

Proof of Proposition 1 The proof is a direct adaptation to my setting of arguments by Maskin (1999) and Moore and Repullo (1990), and is thus relegated to the supplemental appendix. ∎

Proof of Proposition 2 The sufficient condition is obvious. For the necessary condition, single-valuedness follows at once from the fact that individual choice functions are single-valued. Next, suppose that f is implementable in dominant strategies via the mechanism $((M_i)_{i \in I}, \mu)$. Let $m^* : \Theta \to M$ be a profile of strategies such that $m_i^*(\theta)$ is a dominant strategy at θ, and $m_i^*(\cdot)$ varies only with θ_i (such a profile can always be constructed since a dominant action at a state θ remains dominant for individual i at any θ' such that $\theta_i' = \theta_i$, given private values). Let then $\mathcal{M}_i$ be the set of messages that do not belong to the image of the function $m_i^*(\cdot)$. For each $i \in I$, let $g_i : \Theta_i \cup \mathcal{M}_i \to M_i$ be the function defined as follows: $g_i(\mathfrak{m}_i) = m_i^*(\mathfrak{m}_i)$ if $\mathfrak{m}_i \in \Theta_i$ and $g_i(\mathfrak{m}_i) = \mathfrak{m}_i$ if $\mathfrak{m}_i \in \mathcal{M}_i$. Let then $\mu^* : \times_{i \in I}(\Theta_i \cup \mathcal{M}_i) \to X$ be the function defined as follows: $\mu^*(\mathfrak{m}) = \mu((g_i(\mathfrak{m}_i))_{i \in I})$. It is easy to check that truth-telling is a dominant strategy in every type profile for this new mechanism. ∎

Proof of Proposition 3 For the necessary condition, consider a 'direct' mechanism as derived in Proposition 2, and let μ denote its outcome function. Define $X_i(\theta_{-i}) = \{\mu(m_i, \theta_{-i}) | m_i \in \Theta_i \cup \mathcal{M}_i\}$, for each i, θ. Property P then follows at once from the fact that truth-telling is a dominant strategy. As for the sufficient condition, the fact that f satisfies Property P for $(X_i^*)_{i \in I}$ implies that truthtelling is a dominant strategy in the direct mechanism where individuals report their types and the outcome function coincides with f. ∎

Proof of Proposition 4 Let's start by showing that F^{BRE} is not Nash implementable. Suppose that $X = \{x_1, \ldots, x_K\}$. Consider a state $\theta = (\theta_i)_{i \in I}$ where, for each $i \neq 1$, $C_i(Y, \theta_i)$ is the option in Y that has the lowest index (satisfies IIA). Suppose that $C_1(\cdot, \theta_1)$ follows the same pattern, except that $C_1(\{x, x_K\}, \theta_1) = x_K$ instead of x, for each $x \in X \setminus \{x_K\}$ (violates IIA). Hence $x_K \in F^{BRE}(\theta)$. Proposition 1' in the supplemental Appendix extends

Proposition 1 to the case of multi-valued SCRs. By definition of θ, the first part of the necessary condition in Proposition 1' can be satisfied only if $X_i(x_K, \theta) = \{x_K\}$, for all $i \neq 1$, and $X_1(x_K, \theta) = \{x_K\}$, or $\{x_K, x\}$, for some $x \in X \setminus \{x_K\}$. If $X_1(x_K, \theta) = \{x_K\}$, then the second part of the necessary condition implies that $x_K \in F^{BRE}(\theta)$, for all θ, which is clearly not true. In the case where $X_1(x_K, \theta) = \{x_K, x\}$, for some $x \in X \setminus \{x_K\}$, consider an option y different from both x and x_K, and the state (θ'_1, θ_{-1}), where θ'_1 differs from θ_1 only in that $C_1(\{x_K, y\}, \theta'_1) = y$ instead of x_K. The second part of the necessary condition in Proposition 1' implies that x_K must belong to F^{BRE} if it were Nash implementable, but it doesn't, as y unambiguously dominates x_K.

I now show that F^{Eff} is Nash implementable on any Θ. By definition of F^{Eff}, one can associate to any $x \in F^{Eff}(\theta)$ a collection $(Y_i^{x,\theta})_{i \in I}$ of subsets of X such that $C_i(Y_i^{x,\theta}, \theta) = x$, for each $i \in I$, and $X = \cup_{i \in I} Y_i^{x,\theta}$. Let's then check the sufficient conditions in Proposition 1'. Take $X_i(x, \theta) = Y_i^{x,\theta}$. The two parts of the necessary condition follow at once by definition of F^{Eff}. As for the additional condition needed for sufficiency, it is trivially satisfied by F^{Eff} since $F^{Eff}(\theta)$ includes by definition any x such that $x = C_i(X, \theta)$, for some $i \in I$.

Consider a state θ, and $x \in F^{Eff}(\theta)$. The condition $C_i(Y_i, \theta) = x$ implies that none of the elements in $Y_i \setminus \{x\}$ unambiguously dominates x for i. The condition $X = \cup_{i \in I} Y_i$ implies that it is impossible to find an alternative that would be unambiguously preferred to x by all individuals. In other words, x is BR-efficient, and I have proved that $F^{Eff} \subseteq F^{BRE}$.[17]

Finally, if θ is such that $C_i(\cdot, \theta)$ is rational, for each $i \in I$, and x is Pareto efficient for the associated revealed preferences, then define Y_i as the lower contour set of x according to the revealed preference, for each i. By definition, $C_i(\cdot, \theta)(Y_i) = x$, for all i. If there is $x' \in X$ such that $x' \in X \setminus Y_i$, for each $i \in I$, then x' is revealed preferred to x for all i, thereby contradicting the efficiency of x. This shows that $X = \cup_{i \in I} Y_i$, and hence F^{Eff} contains all the Pareto efficient options. The other inclusion, namely that $F^{Eff}(\theta)$ is included

---

[17]Going back to the example presented in the first paragraph of the proof, notice how $x_K \in F^{BRE}(\theta) \setminus F^{Eff}(\theta)$.

in the Pareto set, follows from the previous paragraph, since F^{BRE} coincides with the Pareto set when individuals are rational. ∎

Proof of Proposition 5   The definition of the core is closely related to that of efficiency, applying a similar argument to all coalitions instead of just the grand coalition. The proof thus shares some similarity with that of Proposition 4, and is relegated to the supplemental appendix. ∎

Proof of Proposition 6   I provide the proof only for the case of Nash implementability. The proof for dominant strategy implementation is very similar and thus left to the reader.

Let x, y, z be any three elements in the range of f, let Θ_i^{xyz} be the set of all six rational strict orderings on $\{x, y, z\}$, and let $\alpha_i^{xyz} : \Theta_i^{xyz} \to \Theta_i$ be a function that associates a type $\theta_i \in \Theta_i$ to each $\succ_i \in \Theta_i^{xyz}$ such that x is most preferred according to $C_i(\cdot, \theta_i)$, y is second most preferred, and z is third most preferred. Next consider the SCR $\phi^{xyz} : \Theta^{xyz} \to X$ defined as follows:

$$\phi^{xyz}(\succ) = f(\alpha^{xyz}(\succ)),$$

for each $\succ \in \Theta^{xyz}$ (with the convention $\alpha^{xyz}(\succ) := (\alpha_i^{xyz}(\succ_i))_{i \in I}$). The proof of this lemma now proceeds in various steps.

Step 1 $Range(\phi^{xyz}) = \{x, y, z\}$.

Proof: For each $a \in \{x, y, z\}$, consider the profile of strict orderings $\succ$ on $\{x, y, z\}$ such that a is top-ranked by all individuals. Hence a is most-preferred by all individuals at $\alpha^{xyz}(\succ)$, and BR-efficiency implies that $\phi^{xyz}(\succ) = f(\alpha^{xyz}(\succ)) = a$. Hence $\{x, y, z\} \subseteq Range(\phi^{xyz})$. I prove the opposite inclusion by contradiction. By definition of most preferred, second most preferred, and third most preferred elements, any element in $\{x, y, z\}$ unambiguously dominates any element in $X \setminus \{x, y, z\}$ at type profiles in the image of α^{xyz}. The definition of ϕ^{xyz} and the fact that f is BR-efficient then imply that $Range(\phi^{xyz}) \subseteq \{x, y, z\}$ as well, and hence $Range(\phi^{xyz}) = \{x, y, z\}$. □

Step 2 ϕ^{xyz} is Maskin monotonic.

Proof: Suppose that $\phi^{xyz}(\succ) = a \in \{x, y, z\}$, and that a's rank increases for each individual when moving from $\succ$ to $\succ'$. $a = \phi^{xyz}(\succ) = f(\alpha^{xyz}(\succ))$ and the first part of Property M imply that $C_i(X_i(\alpha^{xyz}(\succ)), \alpha_i^{xyz}(\succ_i)) = a$, for

31

each $i \in I$. The definition of $\succ'$ implies that $C_i(X_i(\alpha^{xyz}(\succ)), \alpha_i^{xyz}(\succ'_i)) = a$, for each $i \in I$. The second part of property M implies that $f(\alpha^{xyz}(\succ')) = a$, and hence $\phi^{xyz}(\succ') = a$, as desired. $\square$

Step 3 *There exists a unique j such that, for all a, b, c in A and all $\succ \in \Theta^{abc}$, $\phi^{abc}(\succ)$ is top ranked for $\succ_j$.*

Proof: By the previous step and the usual result on the rational domain, for each x, y, z, there exists a unique j such that $\phi^{xyz}(\succ) = top(\succ_j)$, for each $\succ \in \Theta^{xyz}$. I just need to prove that j does not vary with x, y, z. It is obviously sufficient to show that x, y, z and x, y, z' lead to the same j. Suppose, on the contrary, that the first triplet leads to j, while the second leads to j', with $j \neq j'$. Consider then $\succ \in \Theta^{xyz}$ such that x is top ranked according to $\succ_j$, y is top ranked according to $\succ_i$, and x is second best according to $\succ_i$, for each $i \in I \setminus \{j\}$. By definition of j, I must have $\phi^{xyz}(\succ) = x$. The first part of property M implies that $x \in X_i(\alpha^{xyz}(\succ))$ and $y \notin X_i(\alpha^{xyz}(\succ))$, for each $i \in I \setminus \{j\}$. Let $\succ' \in \Theta^{xyz'}$ be derived from $\succ$ by replacing z by z'. I have that $C_i(X_i(\alpha^{xyz}(\succ)), \alpha_i^{xyz'}(\succ'_i)) = x$, for all $i \in I$, and hence $\phi^{xyz'}(\succ') = x$, by the second part of Property M and the definition of $\phi^{xyz'}$. This implies a contradiction with the definition of j', as desired. $\square$

Step 4 *There exists $j \in I$ such that $f(\theta) = a$, for each $a \in Range(f)$ and each $\theta \in \Theta$ such that a is most preferred according to $C_j(\cdot, \theta_j)$.*

Proof: Let j be the individual as identified in the previous step, and let a and θ be as in the statement. Suppose, to the contrary of what I want to prove, that $f(\theta) = b \neq a$. The first part of Property M implies that $a \notin X_j(\theta)$. Let c be a third element in the range of f, and let $\succ \in \Theta^{abc}$ be such that a is top ranked according to $\succ_j$, b is second best according to $\succ_j$, and b is top ranked according to $\succ_i$, for each $i \in I \setminus \{j\}$. The second part of Property M implies that $\phi(\succ) = f(\alpha(\succ)) = b$, thereby contradicting the definition of j. $\square$

Step 5 *There exists $j \in I$ and $X_j : \Theta \to P(X)$ such that $Range(f) \subseteq X_j(\theta)$ and $f(\theta) = C_j(X_j(\theta), \theta_j)$, for each $\theta \in \Theta$.*

Proof: Pick j as identified in the previous step. The existence of the X_j follows from Property M. All what needs to be proved is that $Range(f) \subseteq X_j(\theta)$, for all θ. Suppose, on the contrary, that $x \in Range(f) \setminus X_j(\theta)$, for some $\theta \in \Theta$.

So $f(\theta) \neq x$. Part 2 of Property M then implies that $f(\theta'_j, \theta_{-j}) = f(\theta)$, for any θ'_j such that x is most preferred and $f(\theta)$ is second most preferred according to $C_j(\cdot, \theta_j)$, a contradiction with the previous step. ∎

Proof of Proposition 7   The proof for Nash implementability is provided in the supplemental appendix. As for dominant strategies, consider the following mechanism: $M_i = \Theta_i \cup \{m^*\}$, $M_j = \Theta_j$, $M_k = \emptyset$, for all $k \neq i, j$, $\mu(m_i, m_j) = f_{ij}(m_i, m_j)$ if $(m_i, m_j) \in \Theta_i \times \Theta_j$, and $\mu(m^*, m_j) = \gamma(C_j(Z, m_j))$.

I simply have to check that truth-telling is a dominant strategy for each participant. If individual j reports m_j, then the set of outcomes that individual i can generate by varying his report is $Z \cup \{\gamma(C_j(Z, m_j))\}$. If i's type is $\theta_i \in A_i$, then he wants to pick the outcome $C_j(Z, m_j)$, which coincides with $f_{ij}(\theta_i, m_j) = \mu(\theta_i, m_j)$, and hence truth-telling is a dominant strategy. If i's type is $\theta_i \in \Theta_i \setminus A_i$, then he wants to pick the outcome $C_i(Z, \theta_i)$, which coincides with $f_{ij}(\theta_i, m_j) = \mu(\theta_i, m_j)$, and hence truth-telling is a dominant strategy. Consider now individual j. If individual i reports $m_i \in A_i$, then the set of outcomes that individual j can generate by varying his report is Z. If j's type is θ_j, then he wants to pick the outcome $C_j(Z, \theta_j)$, which coincides with $f_{ij}(m_i, \theta_j) = \mu(m_i, \theta_j)$, and hence truth-telling is a dominant strategy. If individual i reports $m_i \in \Theta_i \setminus A_i$, then the set of outcomes that individual j can generate by varying his report is $\{C_i(Z, \theta_i)\}$. Truth-telling is thus trivially a dominant strategy as well, for all types of individual j. Finally, if individual i reports m^*, then the set of outcomes that individual j can generate by varying his report is $\gamma(Z)$. If j's type is θ_j, then he wants to pick the outcome $\gamma(C_j(Z, \theta_j))$, which coincides with $f_{ij}(m^*, \theta_j) = \mu(m^*, \theta_j)$, and hence truth-telling is a dominant strategy. Hence my mechanism does indeed implement f_{ij} in dominant strategies. ∎

Proof of Proposition 8   The sufficient condition, namely that such f's are Nash implementable, is straighforward to check. I thus focus on the necessary condition. Let $\rho_i : \Theta_i^* \to \Theta_i$ be the function that associates to each rational type θ_i^* the (unique) type θ_i such that $C_i(S, \theta_i) = \arg\max_{\succ(\theta_i^*)} A_i(S)$, for all $S \in P(X)$ (cf. equation (3)). Define $\rho : \Theta^* \to \Theta$ and $\phi : \Theta^* \to X$ as follows:

$\rho(\theta^*) = (\rho_i(\theta_i^*))_{i \in I}$ and $\phi(\theta^*) = f(\rho(\theta^*))$, for each $\theta^* \in \Theta^*$. I now prove that ϕ is Maskin monotonic. Consider thus $\theta^*, \hat{\theta}^* \in \Theta^*$ such that

$$LC_i(\phi(\theta^*), \theta_i^*) \subseteq LC_i(\phi(\theta^*), \hat{\theta}_i^*). \tag{5}$$

If f is Nash implementable, then it satisfies property M, thanks to Proposition 1. Hence $f(\rho(\theta^*)) = C_i(X_i(\rho(\theta^*)), \rho_i(\theta_i^*))$, or $f(\rho(\theta^*)) \in A_i(X_i(\rho(\theta^*)))$ and $A_i(X_i(\rho(\theta^*))) \subseteq LC_i(\phi(\theta^*), \theta_i^*)$. Combining this with condition (5), it must be that $f(\rho(\theta^*))$ is the maximal element in $A_i(X_i(\rho(\theta^*)))$ according to $\succ(\hat{\theta}_i^*)$ as well. Part 2. of Property M thus implies that $f(\rho(\hat{\theta}^*)) = f(\rho(\theta^*))$, or $\phi(\hat{\theta}^*) = \phi(\theta^*)$, as desired. Applying a standard result on the rational domain, I conclude that there exists $j \in I$ such that $\phi(\theta^*) = \arg\max_{\succ(\theta_j^*)}[Range(\phi)]$, for each $\theta^* \in \Theta^*$. Notice that $Range(\phi) \subseteq Range(f)$, by definition of ϕ. Condition (2) implies the converse inclusion, and hence

$$\phi(\theta^*) = \arg\max_{\succ(\theta_j^*)}[Range(f)], \tag{6}$$

for each $\theta^* \in \Theta^*$.

Property M implies that $f(\theta) = C_j(X_j(\theta), \theta_j) = \arg\max_{\succ_{\theta_j}} A_j(X_j(\theta))$, for each $\theta \in \Theta^*$. I now prove that $Range(f) \subseteq A_j(X_j(\theta))$, for each θ. Suppose, to the contrary of what I want to prove, that there exist $\theta \in \Theta$ and $x \in Range(f) \setminus A_j(X_j(\theta))$. Let $\theta^* \in \Theta^*$ be such that $\rho(\theta^*) = \theta$ (such a θ^* exists, thanks to condition (2)). Consider then $\hat{\theta}_j^* \in \Theta_j^*$ obtained from θ_j by bringing x to the top, while keeping all the other comparisons unchanged. The second part of property M implies that $\phi(\hat{\theta}_j^*, \theta_{-j}^*) = f(\rho_j(\hat{\theta}_j^*), \theta_{-j}) = f(\theta)$ (because $x \notin A_j(X_j(\theta))$), which then contradicts the fact that $\phi(\hat{\theta}_j^*, \theta_{-j}^*) = \arg\max_{\succ(\hat{\theta}_j^*)} Range(f)$. In order to avoid this contradiction, it must thus be the case that $Range(f) \subseteq A_j(X_j(\theta))$, for each θ. Consider now $\bar{\theta}^* \in \Theta^*$ such that the elements in the range of f are ranked below any other element of X according to $\succ(\bar{\theta}_j^*)$. If there exists $x \in A_j(X_j(\rho(\bar{\theta}^*))) \setminus Range(f)$, then one would reach a contradiction with the fact that $f(\rho(\bar{\theta}^*)) = \arg\max_{\succ(\rho(\bar{\theta}_j^*))} A_j(X_j(\rho(\bar{\theta}^*)))$. Hence it must be that $A_j(X_j(\rho(\bar{\theta}^*))) = Range(f)$. The result then follows by choosing $S = X_j(\rho(\bar{\theta}^*))$. ∎

Proof of Proposition 9   I start by showing that f^l is Nash implementable. This will follow from Proposition 1 after checking that f^l satisfies Property $\mathfrak{M}$. Let $i \in I$ and $\theta \in \Theta$. Define $X_i(\theta) = X$ if $C_i(X, \theta_i) = f^l(\theta)$, and $= \{x_1, \ldots, f^l(\theta)\}$ if $C_i(X, \theta_i) \neq f^l(\theta)$.

I start by checking the first condition in Property M. If $C_i(X, \theta_i) = f^l(\theta)$, then $X_i(\theta) = X$ and $C_i(X_i(\theta), \theta_i) = f^l(\theta)$. If $C_i(X, \theta_i) \neq f^l(\theta)$, then $C_i(X, \theta_i)$ comes after $f^l(\theta)$ in the list, and hence $C_i(X_i(\theta), \theta_i)$ is the last element of $X_i(\theta)$, which is $f^l(\theta)$, as desired. Let's now check the second condition in Property M. Let θ, θ' be such that $C_i(X_i(\theta), \theta_i') = f^l(\theta)$, for each $i \in I$. Notice that $f^l(\theta')$ cannot come strictly after $f^l(\theta)$ in the list, since $X_i(\theta) = X$, for at least one individual i. Notice that it cannot come strictly before either, as this would require $C_i(X, \theta_i')$ to fall strictly before $f^l(\theta)$, for at least one individual i, which would contradict $C_i(X_i(\theta), \theta_i') = f^l(\theta)$. Hence it must be that $f^l(\theta') = f^l(\theta)$.

I can conclude the proof of Nash implementability by checking the additional condition that distinguishes Property $\mathfrak{M}$ from M. Let $j \in I$, $\theta, \theta' \in \Theta$, and $x \in X$ be such that $C_i(X, \theta_i') = x$, for each $i \in I \setminus \{j\}$, and $C_j(X_j(\theta), \theta_j') = x$ or $C_j(X, \theta_j') = x$. If $C_j(X, \theta_j') = x$, then $f^l(\theta') = x$, by definition of f^l. If $C_j(X_j(\theta), \theta_j') = x$, then $C_j(X, \theta_j')$ does not come before x in the list. Combining that with the other conditions on the choice for each $i \neq j$, I conclude that $f(\theta') = x_{k^l(\theta')} = x$, as desired.

I now check that f^l is implementable in dominant strategies. For each i and each θ, let $k_{-i}^l(\theta) = \min\{k \in \{1, \ldots, |X|\} | (\exists j \in I \setminus \{i\}) : x_k = C_j(X, \theta_j)\}$. Observe that $k_{-i}^l(\theta)$ is independent of θ_i. Hence $\{x_k \in X | k \leq k_{-i}^l(\theta)\}$ is also independent of θ_i, and will thus be denoted $X_i(\theta_{-i})$. Let's check that $f^l(\theta) = C_i(X_i(\theta_{-i}), \theta_i)$, for all θ. Suppose first that $k^l(\theta) = k_{-i}^l(\theta)$. Then $C_i(X, \theta_i)$ does not come before $x_{k_{-i}^l(\theta)}$ in the list, and $C_i(X_i(\theta_{-i}), \theta_i) = x_{k_{-i}^l(\theta)}$, given the assumption on choice functions. Hence $C(X_i(\theta_{-i}), \theta_i) = x_{k^l(\theta)} = f^l(\theta)$, as desired. Suppose next that $k^l(\theta) < k_{-i}^l(\theta)$. Hence $C_i(X, \theta_i)$ comes before $x_{k_{-i}^l(\theta)}$ in the list, and $C_i(X_i(\theta_{-i}), \theta_i) = C_i(X, \theta_i) = x_{k^l(\theta)}$, given the assumption on choice functions. Again, $C_i(X_i(\theta_{-i}), \theta_i) = f^l(\theta)$, as desired.

The result from the previous paragraph also implies that $\{f^l(\theta_i, \theta_{-i}) | \theta_i \in$

$\Theta_i\} = \{C_i(X_i(\theta_{-i}), \theta_i)|\theta_i \in \Theta_i\} = X_i(\theta_{-i})$. The sufficient condition from Proposition 3 applies, and f^l is thus implementable in dominant strategies. ∎

Proof of Proposition 10 Let Θ_i^* be a set encoding all the choice functions associated to the maximization of a preference ordering that is single-peaked on the list. Notice that applying the "stop when you start to decline" procedure to a single-peaked preference delivers the same choice function as by maximizing it. Hence I may think of Θ_i^* as a subset of Θ_D and thus also of Θ. The simple mechanism will implement f in dominant strategies on Θ^*. By Moulin (1980), there must exist $I - 1$ elements of X (interpreted as choices of "phantom" voters) such that f coincides on Θ^* with the median of these points and $\{C_i(X, \theta_i)|i \in I\}$. The unique dominant strategy for type $\theta_i \in \Theta_i^*$ is to send the message $x = C_i(X, \theta_i)$. All the elements of X are thus already used by types in Θ_i^*, and hence, for any set of messages, f must coincide with the median of these messages and the phantom voters' choices. I now conclude the proof by showing that all the phantom voter's choices must fall on the first element of the list. Suppose not. Then it is not difficult to check that one can select messages for $I - 1$ individuals such that, combined with the phantom voters' choices, the set of medians that the missing voter can generate by changing his message constitutes a connected subset Y of the list that does not contain the first element of the list. Consider now an individual i, and a type θ_i^* for that individual such that the first element of the list is best for him, the second one is worst, and each element thereafter in the list is better than the previous. If the individuals other than i send those messages associated to Y, then i's best response is to send a message that coincides or follows the last element of Y. On the other hand, if all the other individuals send the message that coincides with the first element of the list, then i's best response is to send that same message. Hence the simple mechanism has no dominant strategy if phantom voters' choices do not all fall on the first element of the list - a contradiction. Hence all these choices do fall on the first element, and the mechanism must select the left-most message among those that have been sent. Since $\Theta \subseteq \bar{\Theta}$, sending $C_i(X, \theta_i)$ is a best response for each type $\theta_i \in \Theta_i$ and each i, and hence $f = f^l$. ∎

References

Abdulkadiroglu, A. and T. Sonmez, 1998. Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems. *Econometrica* **66**, 689-701.

Aizerman, M. A., and F.T. Aleskerov, 1986. Voting operators in the space of choice functions. *Mathematical Social Sciences* **11**, 201-242.

Ambrus, A., and K. Rozen, 2009. Rationalizing Choice with Multi-Self Models. Cowles Foundation Discussion Paper.

Bade, S., 2008. Housing Problems with Non-Rationalizable Behavior. *Mimeo.*

Baumeister, R. F., and J. Tierney, 2011. Willpower: Rediscovering the Greatest Human Strength. Penguin Press.

Bernhein, B. D., and A. Rangel, 2009. Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics* **124**, 51-104.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue, and M. Rabin, 2003. Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism." University of Pennsylvania Law Review **151**, 1211-1254.

Cherepanov, V., T. Feddersen, and A. Sandroni, 2009. Rationalization. Working paper, Northwestern University.

Corchón, L., 1996. The Theory of Implementation of Socially Optimal Decisions in Economics. St. Martin's Press, New York.

Dasgupta, P. S., P. J. Hammond, and E. Maskin, 1979. The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility. *Review of Economics Studies* **46**, 185-216.

de Clippel, G., and K. Eliaz, 2012. Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects. *Theoretical Economics* **7**, 125-162.

Eliaz, K., 2002. Fault-Tolerant Implementation. *Review of Economic Studies* **69**, 589-610.

Glazer, J., and A. Rubinstein, 2011. A Model of Persuasion with a Boundedly Rational Agent. *Mimeo.*

Huber, J., J. W. Payne, and C. Puto, 1982. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Slimilarity Hypothesis. *Journal of Consumer Research* **9**, 90-98.

Hurwicz, L., 1986. On the Implementation of Social Choice Rules in Irrational Societies, in Heller et al., editors, Essays in Honor of Kenneth J. Arrow,

Vol. I.

Jackson, M. O., 2001. A Crash Course in Implementation Theory. *Social Choice and Welfare* **18**, 655-708.

Kalai, E., E. Muller, and M. A. Satterthwaite, 1979. Social welfare functions when preferences are convex, strictly monotonic, and continuous. *Public Choice* **34**, 87-97.

Kalai, G., A. Rubinstein, and R. Spiegler, 2002. Rationalizing Choice Functions by Multiple Rationales. *Econometrica* **70**, 2481-2488.

Korpela, V., 2012. Implementation without Rationality Assumptions, *Theory and Decision*, forthcoming.

Lipman, B., and W. Pesendorfer, 2011. Temptation. *Mimeo.*

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Ozbay, 2010. When More is Less: Choice by Limited Consideration. *Mimeo.*

Manzini, P., and M. Mariotti, 2007. Sequentially Rationalizable Choice. *American Economic Review* **97**, 1824-1839.

Manzini, P., and M. Mariotti, 2012. Categorize Then Choose: Boundedly Rational Choice and Welfare. *Journal of the European Economic Association*, forthcoming.

Masatlioglu, Y., and E.A. Ok, 2005. Rational Choice with Status Quo Bias, *Journal of Economic Theory* **121**, 1-29.

Masatlioglu, Y., D. Nakajima, and E. Ozbay, 2012. Revealed Attention. *American Economic Review*, forthcoming.

Masatlioglu, Y., D. Nakajima, and E. Ozdenoren, 2011. Revealed Willpower. *Mimeo.*

Maskin, E., 1985. The Theory of Implementation in Nash Equilibrium: A Survey. In Social Goals and Social Organization, L. Hurwicz, D. Schmeidler, and H. Sonnenschein, eds., Cambridge University Press, Cambridge, UK, 173-204.

Maskin, E., 1999. Nash Equilibrium and Welfare Optimality. *Review of Economic Studies* **66**, 23-38.

Maskin, E., and T. Sjöström, 2002. Implementation Theory. In Handbook of Social Choice and Welfare, Vol. I, K. J. Arrow, A. Sen, and K. Suzumura, eds., Elsevier Science, New York, 237-288.

Mookherjee, D., and S. Reichelstein, 1990. Implementation via Augmented Revelation Mechanisms, *Review of Economic Studies* **57**, 453-475.

Moore, J., 1992. Implementation, Contracts and Renegotiation in Environments with Complete Information. In Advances in Economic Theory, 4th

World Congress of the Econometric Society, Vol. I, J. J. Laffont, ed., Cambridge University Press, Cambridge, UK, 182-282.

Moore, J., and R. Repullo, 1990. Nash Implementation: A Full Characterization. *Econometrica* **56**, 1191-1220.

Moulin, H., 1980. On Strategy-Proofness and Single-Peakedness. *Public Choice* **35**, 437-455.

Ok, E., P. Ortoleva, and G. Riella, 2011. Revealed (P)Reference Theory. *Mimeo*.

Ozdenoren, E., S. Salant, and D. Silverman, 2010. Willpower and the Optimal Control of Visceral Urges. *Journal of the European Economic Association*, forthcoming.

Palfrey, T. R., 1992. Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design. In Advances in Economic Theory, 4th World Congress of the Econometric Society, Vol. I, J. J. Laffont, ed., Cambridge University Press, Cambridge, UK, 283-323.

Palfrey, T. R., 2002. Implementation Theory. In Handbook of Game Theory with Economic Applications, Vol. III, R. J. Aumann and S. Hart, eds., Elsevier Science, New York, 2271-2326.

Ray, K. T., 2010. Nash Implementation under Irrational Preferences. *Mimeo*, McDonough School of Business, Georgetown University.

Roberts, J. H., and J. M. Lattin, 1991. Development and Testing of a Model of Consideration Set Composition. *Journal of Marketing Research* **28**, 429-440.

Rubinstein, A., and Y. Salant, 2006. A Model of Choice from Lists. *Theoretical Economics* **1**, 3-17.

Salant, Y., and A. Rubinstein, 2008. (A,f): Choice with Frames. *Review Of Economic Studies* **75**, 1287-1296.

Saran, R., 2011. Menu-Dependent Preferences and Revelation Principle. *Journal of Economic Theory*, forthcoming

Serrano, R., 2004. The Theory of Implementation of Social Choice Rules. *SIAM Review* **46**, 377-414.

Shapley, L., and H. Scarf, 1974. On Cores and Indivisibility. *Journal of Mathematical Economics* **1**, 23-37.

Simon, H. A., 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* **69**, 99-118.

Spiegler, R., 2011. Bounded Rationality and Industrial Organization. *Oxford University Press*.

Thaler R. H., and C. R. Sunstein, 2003. Libertarian Paternalism. *American Economic Review* **93**, 175-179

Thaler R. H., and C. R. Sunstein, 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. *Yale University Press.*

Wright, P., and F. Barbour, 1977. Phased Decision Strategies: Sequels to an Initial Screening. In Studies in Management Sciences, Multiple Criteria Decision Making, ed. Martin K. Starr, and Milan Zeleny, 91-109, Amsterdam: North-Holland.