

# Rationalizing Choice with Multi-Self Models\*

Attila Ambrus<sup>†</sup>  
Harvard

Kareen Rozen<sup>‡</sup>  
Yale

This version: September 2009

First version: April 2008

## Abstract

To facilitate systematic study of multi-self decision making, this paper proposes an axiomatic framework that encompasses a variety of models proposed in economics, psychology, and marketing. We model a decision-maker as a collection of utility functions (selves) and an aggregation rule (a theory of how selves are activated by choice sets) satisfying five simple axioms of social choice. For a broad class of aggregators, we show that with sufficiently many selves the resulting model can rationalize any choice function. We propose a method for counting IIA violations in a choice behavior, and show that the set of behaviors an aggregator can rationalize using  $n$  selves is bounded below by the set of choice functions for which the number of IIA violations is at most a given linear function of  $n$ . We apply our results to study Strotzian choice over menus and household decision-making.

**JEL Codes:** D11, D13, D71

**Keywords:** Multiple selves, IIA violations, rationalizability, complexity

---

\*Previous versions of this paper were distributed under the title “Revealed Conflicting Preferences: Rationalizing Choice with Multi-Self Models.” We are grateful to Eddie Dekel, Drew Fudenberg, John Geanakoplos, Dino Gerardi, Tzachi Gilboa, Jerry Green, Daniel Hojman, Gil Kalai, Bart Lipman, Philippe Mongin, Wolfgang Pesendorfer, Ben Polak, Ariel Rubinstein, Larry Samuelson, Rani Spiegler, and Tomasz Strzalecki for helpful suggestions. We also thank seminar audiences at Brown, Harvard, MIT, Montreal, NYU, UCL, Yale and the North American Summer Meeting of the Econometric Society for insightful comments.

<sup>†</sup>Address: Dept. of Economics, Littauer Center, 1875 Cambridge St, Cambridge, MA 02138. E-mail: [ambrus@fas.harvard.edu](mailto:ambrus@fas.harvard.edu). Home page: <http://www.economics.harvard.edu/faculty/ambrus>.

<sup>‡</sup>Address: Dept. of Economics and the Cowles Foundation for Research in Economics, 30 Hillhouse Ave., New Haven, CT 06511. E-mail: [kareen.rozen@yale.edu](mailto:kareen.rozen@yale.edu). Home page: <http://www.econ.yale.edu/~kr287/>.

# 1 Introduction

The classical model of choice endows the decision-maker (DM) with a single preference relation that she uses to select the best element from any set of alternatives. The single implication of this model is context-independent behavior, or the *Independence of Irrelevant Alternatives* (IIA), which dictates that if an alternative is deemed optimal in a set, it must remain optimal in any subset.<sup>1</sup> Consequently, a growing body of evidence suggesting that behavior is prone to context-dependence has spurred interest in alternative models of decision-making that can facilitate violations of IIA. In particular, since the seminal work of May (1954), many papers have proposed models of multi-self decision making to accommodate such behaviors.<sup>2</sup>

Formal models of multi-self decision-making proposed in the literature include Kalai, Rubinstein and Spiegel (2002), Fudenberg and Levine (2006), Manzini and Mariotti (2007), and Green and Hojman (2009) in economics; Tversky (1969), Shafir, Simonson and Tversky (1993) and Tversky and Simonson (1993) in psychology; and Kivetz, Netzer and Srinivasan (2004) in marketing.<sup>3</sup> Psychologists have long viewed the *multiplicity of self* as a normal feature instead of a sign of pathology; even psychologists who prefer a unitary view of the self accept that “the singular self is a hypothetical construct, an umbrella under which experiences are organized along various dimensions or motivational systems” and which “is fluid in that it shifts in different contexts as various motivations are activated” (Lachmann 1996). The marketing literature interprets selves as different considerations (criteria) that consumers use when evaluating products. The economics literature often views selves as rationales (e.g., Kalai et al. 2002) or manifestations of temptation and self-control processes (e.g., Strotz 1955, Fudenberg and Levine 2006). In general, these models are motivated by the desire to explain a particular empirically observed choice behavior that is inconsistent with rational choice. Some fix the number of selves — as in the dual-self model of Fudenberg and Levine (2006) — while others leave the number unrestricted — as in Kalai et al. (2002).

There has been little effort to connect the various models, or to conduct an analysis of multi-self decision-making using a more systematic approach. In order to facilitate the latter, this paper develops a general framework to examine a DM with multiple selves, when choice sets themselves

---

<sup>1</sup>This also implies transitive choice behavior, which is often violated in experimental settings (e.g., see Tversky (1969) and Lee, Amir and Ariely (2007)).

<sup>2</sup>Another approach, developed in Bernheim and Rangel (2007) and Salant and Rubinstein (2008), allows for context-dependence by considering extended choice situations where behavior can depend on unspecified *ancillary conditions* or *frames*. While information effects can explain some context dependence (Sen (1993), Kochov (2007), Kamenica (2008)), they cannot explain many systematic violations of IIA (Tversky and Simonson (1993)).

<sup>3</sup>An expanded shortlist of the multiple-selves or multiple-utility literature includes Benabou and Pycia (2002), Masatlioglu and Ok (2005), Evren and Ok (2007), and Chatterjee and Krishna (forthcoming). This literature is also related to the application of social choice tools in multi-criteria decision problems, as in Arrow and Raynaud (1986), and is related more generally to the theory of multiattribute utility (see Keeney and Raiffa (1993)).

serve as frames that influence how the preferences of different selves get aggregated. More formally, we model the DM as a collection of utility functions  $U$  (selves) and an *aggregation rule*  $f$  (decision-making method) which combines these utility functions in a possibly context-dependent way. That is, given a choice set  $A$ , and selves  $U$ , an aggregator  $f$  specifies an aggregate utility for every alternative in  $A$ . An aggregator captures a theory of multi-self decision-making. We examine a broad class of aggregators satisfying five simple axioms from social choice theory, and show that many models of multi-self decision-making proposed in the existing literature can be formally translated into an aggregator in our framework.

One important feature of our framework is that the aggregation can potentially depend on cardinal information in the selves' utilities. The motivation for this is twofold. First, many existing models of multi-self decision-making do make use of cardinal information embedded in different selves' utility functions. Second, in intrapersonal decision-making, the intensities of preferences of different motivations can play an important role. As opposed to interpersonal preference aggregation, these intensities are comparable by the decision-maker. Indeed, such cardinal comparisons are assumed in expected utility theory: a DM trades off utility across possible states. To motivate such comparisons in our framework with multiple rationales, suppose a person choosing where to live cares about his children as well as proximity to work. One possible home is adjacent to his workplace in the city but the school is considered unsafe; the other home is in a suburb which would be a short commute to work but the school is safer. Without cardinality, it is difficult to argue that it is much more important for the children to be in a safe school than it is to have a short commute to work. On the other hand, it is plausible to assume that the person is willing to trade a small enough degree of safety for a substantially reduced commuting time.

A second important feature of our model is the possibility of compromise among selves. Psychologists believe that a fluid form of compromise among selves is necessary for healthy behavior.<sup>4</sup> As opposed to the models provided in Kalai et al. (2002) and Cherepanov, Feddersen and Sandroni (2008), but in accordance with models proposed in Tversky (1969), Tversky and Kahneman (1991), Kivetz et al. (2004), Fudenberg and Levine (2006), Green and Hojman (2009) and others, all of the selves in our framework are “active” for every possible choice set. However, the weights allocated to different selves by the aggregator can depend on the choice set. This means that the model can capture behavior as in Fudenberg and Levine (2006), where a long-run self must exert more costly self control when more appealing options are available to a short-run self; or Shafir et al. (1993), where rationales for purchasing are activated by the set of available products.

Within this framework, formalized in [Section 2](#), we investigate the set of behaviors that a specific

---

<sup>4</sup>This is as opposed to disassociated selves (i.e., overly autonomous selves), or a high self-concept differentiation (a lack of interrelatedness of selves across contexts) both of which are connected to pathological or unhealthy behavior; see Power (2007), Donahue, Robins, Roberts and John (1993), and Mitchell (1993).

model of multi-self decision-making, as captured by a given aggregation rule  $f$ , can *rationalize* (explain). We address this question both with a fixed number of selves, as well as with no *a priori* restriction. Since our results apply for a broad class of aggregators, this paper provides a meta-analysis of various multi-self models proposed in the literature.

Formally, the DM’s behavior is described by a choice function  $c$  that specifies the alternative she selects from each subset of the grand set of alternatives  $X$ . For a given model of aggregation  $f$ , the DM’s choice behavior is *rationalized* by a finite collection of selves  $U$  if she selects the unique maximizer of aggregate utility  $f \circ U$  from every choice set. The DM’s choice behavior need not satisfy IIA. To characterize the extent to which a choice behavior deviates from rationality, we develop in [Section 3](#) a method of counting IIA violations.

For some aggregators, it is straightforward to determine the set of choice functions that can be rationalized. For example, if the DM’s method of aggregating the utilities of her selves is simple utilitarianism, then the set of choice functions is exactly the set of rational choice functions — regardless of the number of selves. But what if, in analogy to models of relative utilitarianism (e.g., Karni 1998), each self’s utility is normalized by her range of utilities over the choice set? Or if the aggregator is the “normalized contextual concavity model” proposed in Kivetz et al. (2004),

$$\sum_{u \in U} (\max_{a' \in A} u(a') - \min_{a' \in A} u(a')) \cdot \left[ \frac{u(a) - \min_{a' \in A} u(a')}{\max_{a' \in A} u(a') - \min_{a' \in A} u(a')} \right]^\rho?$$

Our main results, in [Section 4](#), establish that for a large class of multi-self models, including various models proposed in the existing literature, if there is no restriction on the number of selves then the model can rationalize any choice function. Indeed, for a subclass of aggregators, whenever two simple types of irrational behavior can be rationalized on a triple of alternatives, the aggregator can rationalize any behavior over any set of alternatives. Hence, without restricting the number of selves, a multi-self model might not have testable restrictions on behavior. The lesson to draw from this is that to offer a refutable theory of multi-self decision-making, it is not enough to impose a concrete method of aggregating different selves’ utilities; it is also important to fix the number of selves *a priori* (e.g., as in a “dual-self” model) in order to restrict the set of behaviors that the model can rationalize.

How the number of selves affects the predictions of a given model relates to our second main result. For a fixed number of selves, we use our measure of IIA violations to characterize a lower bound on the set of choice functions that can be rationalized; and show there is a linear relationship between the number of selves and the number of IIA violations in the bound. The bounds we obtain on rationalizable behavior are tightened when the IIA violations in a choice function are related to each other in a systematic manner.

We apply our framework and results to two different contexts. In [Section 5](#) we propose a generalized Strotzian model: a DM chooses a menu in anticipation of her future choice from that menu, which arises as a compromise among multiple motivations. In this setting behavior that is interpreted differently by the literature on choice over menus can arise from “anticipated” IIA violations. In [Section 6](#) we analyze collective household choice, extending our analysis to incomplete choice functions, such as demand functions. Our results complement those of Browning and Chiappori (1998) and Chiappori and Ekeland (2006) in this context.

Our results draw a connection between the complexity of a rationalization and the extent to which the choice behavior in question deviates from rationality, as measured by the number of IIA violations.<sup>5</sup> Our framework and results differ from Kalai et al. (2002), who were the first to address whether a given choice behavior can be rationalized, and to examine the required complexity of a rationalization as a function of the number of alternatives available. They propose that a collection of strict preference relations rationalizes a choice function if the choice from each set is optimal for at least one of the preference relations. In this view, each self serves as a dictator for some subset of choices. In contrast, in our framework it can happen that the choice is not the most preferred alternative of any of the selves, but the *best compromise*, in the sense that it maximizes aggregate utility.

There are also several recent contributions to the literature on multi-self decision-making, which mostly focus on a different set of questions than we do. Of these, the most related is Green and Hojman (2009), who also study a class of aggregation methods. Because they model a DM as a probability distribution over all possible ordinal preference rankings, their framework is difficult to compare to models of multi-self decision-making with a discrete number of cardinal selves, but is related to models in the voting literature (e.g., Saari 1999). Extending results from that literature, they show that if choice is determined by a voting rule satisfying a monotonicity property, then their model can explain any choice behavior.<sup>6</sup> The rest of the paper focuses on welfare analysis. Bernheim and Rangel (2007) and Chambers and Hayashi (2008) also focus on welfare analysis given choices contradicting rational decision-making. Other related work includes Manzini and Mariotti (2007), Masatlioglu and Nakajima (2007) and Cherepanov et al. (2008), who consider sequential application of multiple rationales to eliminate alternatives, a process they show can rationalize certain choice functions. Finally, Fudenberg and Levine (2006) consider a dual-self model of dynamic choice, where the two selves’ utilities are aggregated in a menu-dependent way.<sup>7</sup>

---

<sup>5</sup>Measuring the complexity of a rationalization by the number of selves is akin to measuring the complexity of an automata by the number of states (e.g., see Salant (2007) in the context of decision-making).

<sup>6</sup>Our result on rationalization is independent of this monotonicity theorem.

<sup>7</sup>See also Chatterjee and Krishna (forthcoming) for a model of dual-self decision-making.

## 2 A framework for rationalizing choice

### 2.1 Main concepts and definitions

Suppose that we observe a DM’s choice behavior on a finite set of alternatives  $X$ . Denote by  $P(X)$  the set of nonempty subsets of  $X$ . The DM’s *choice function*  $c : P(X) \rightarrow X$  identifies the alternative  $c(A) \in A$  that she chooses from each  $A \in P(X)$ . A *rationalization* of the DM’s choice function consists of two components, a collection of *selves*  $U$  and an *aggregator*  $f$  that combines the utilities of different selves, in a possibly menu-dependent way, into an aggregated utility function. The DM’s selves represent her conflicting motivations or priorities. The aggregator corresponds to the DM’s method of “sorting out” her priorities to come to a decision. In order to simplify notation and terminology, in the main text we define a simplified framework in which the aggregator treats all selves symmetrically. However, in order for our framework to encompass as many of the multi-self models proposed in the existing literature as possible, in [Appendix A](#) we permit selves to have “types” and extend the construction and our main results to asymmetric aggregators that treat selves differently according to their type.

Formally, given a grand set of alternatives  $X$ , a self (a.k.a. *reason*, *rationale*) is a utility function  $u : X \rightarrow \mathbb{R}$ . Hence, each self is an element of  $\mathbb{R}^X$ , and  $u(x)$  is the utility level that self  $u$  allocates to  $x \in X$ .<sup>8</sup> A *collection of selves*  $U$  is an unordered list of selves.<sup>9</sup> By definition of an unordered list, a collection of selves can have multiple selves with the same utility function  $u$ , and there is no order hierarchy defined over elements of the list. Formally, for a given grand set of alternatives  $X$ , a collection of selves  $U$  is an element of  $\mathcal{U}(X) = \cup_{n=1}^{\infty} \mathcal{U}^n(X)$ , where  $\mathcal{U}^n(X)$  is the set of all unordered lists of selves over  $X$  that contain  $n$  elements. To denote the number of selves in a particular collection  $U$ , we use the notation  $|U|$ , or simply  $n$  when no confusion would arise.

An aggregator  $f$  specifies an aggregate utility for every alternative  $a$  in every choice set  $A$ , given any (finite) grand set of alternatives  $X$  and any collection of selves  $U$  defined over these alternatives. Formally, the domain over which  $f$  is defined is  $\{a, A, X, U \mid X \in \mathcal{X}, U \in \mathcal{U}(X), A \in P(X), a \in A\}$ , where  $\mathcal{X}$  is the set of conceivable finite grand sets of alternatives. Since the choice set  $A$  is one of the arguments of the function,  $f$  aggregates the utilities of the selves in a possibly context-dependent way.<sup>10</sup> The grand set of alternatives  $X$  appears as an argument of the aggregator, not only because the evaluation of an alternative  $a \in A$  could potentially depend on alternatives outside the choice

---

<sup>8</sup>Though aggregation in our framework is cardinal, the model has the “ordinal” feature that there can be many “equivalent” representations of an aggregator in this context. In particular, if  $f$  rationalizes the choice function  $c$  using the selves  $U$ , then so does any increasing transformation of  $f$ . Similarly, if  $f$  rationalizes  $c$  using the selves  $U$ , then  $f \circ h^{-1}$  rationalizes  $c$  using the selves  $h \circ U$ , where  $h : \mathbb{R} \rightarrow \mathbb{R}$  is invertible on the appropriate domain. That is, given any representation  $U$  and  $f$ , one can obtain an equivalent representation by applying a monotone transformation of utilities in  $U$ , if a corresponding transformation is applied to the aggregation function  $f$  as well.

<sup>9</sup>In combinatorics this object is also referred to as a multiset.

<sup>10</sup>We could permit aggregators with restricted domains: let  $\hat{\mathbb{R}}^X$  be a convex subset of  $\mathbb{R}^X$  and let  $\mathcal{U}^n = \times_{i=1}^n \hat{\mathbb{R}}^X$ .

set  $A$ , but also because this enables a “comparative static”: we study how the number of selves needed to rationalize a choice rule depends on the size of  $X$ .

Given an aggregator, we say that a collection of selves rationalizes a choice function if from every choice set the alternative that maximizes the aggregated utility is the one selected by the choice function.

**Definition 1.** *Choice function  $c(\cdot)$  on  $X$  is rationalized by the aggregator  $f$  if there exists a finite collection of selves  $U \in \mathcal{U}(X)$  such that for every  $A \in P(X)$ ,  $c(A) = \arg \max_{a \in A} f(a, A, X, U)$ .*

## 2.2 Basic axioms of aggregation

We are interested in examining theories of aggregation that are in line with the underlying selves’ preferences. For this reason, for the rest of the paper we restrict attention to aggregators satisfying the following properties, most of which are familiar from the theory of social choice. These properties are also satisfied by several previously proposed models of multi-self decision-making. As we will argue below, imposing these properties is a natural requirement if the aggregation of utilities is cardinal and the framing effect of a choice set operates only through the utility levels of alternatives for different selves.

To state these properties, define a collection of selves  $U \in \mathcal{U}(X)$  to be  $\delta$ -indifferent with respect to  $X$  if  $\max_{a, b \in A, A \subseteq X} |f(a, A, X, u) - f(b, A, X, u)| < \delta$  for every  $u \in U$ . Also, for any two collections of selves  $U, U' \in \mathcal{U}(X)$ , denote by  $(U, U')$  the combined collection  $(u_1, u_2, \dots, u_{|U|}, u'_1, u'_2, \dots, u'_{|U'|}) \in \mathcal{U}(X)$ .

**P1** (Neutrality). *For any permutation  $\pi : X \rightarrow X$ ,  $f(\pi(a), \pi(A), X, U \circ \pi^{-1}) = f(a, A, X, U)$ .*

**P2** (Single-self respect). *For any  $u \in \mathbb{R}^X$ ,  $u(a) \geq u(b)$  if and only if  $f(a, A, X, u) \geq f(b, A, X, u)$ .*

**P3** (Separability). *If  $f(a, A, X, U) \geq f(b, A, X, U)$  and  $f(a, A, X, \hat{U}) \geq f(b, A, X, \hat{U})$  then  $f(a, A, X, (U, \hat{U})) \geq f(b, A, X, (U, \hat{U}))$ , with strict inequality if one of the above is strict.*

**P4** (Continuity at indifferent selves). *If  $f(a, A, X, U) > f(b, A, X, U)$  then for any  $k \in \mathbb{Z}_+$  there is  $\delta_k > 0$  such that  $f(a, A, X, (U, U')) > f(b, A, X, (U, U'))$  for any  $\delta_k$ -indifferent  $U' \in \mathcal{U}^k(X)$ .*

**P5** (Duplication). *If  $U(a) = U(\hat{a})$  then  $f(b, A \cup \{a\}, X, U) = f(b, A \cup \{\hat{a}\}, X, U)$  for every  $b \in A$ .*

Neutrality implies that the names of alternatives do not affect their ranking (only utilities affect rankings). Single-self respect is a minimal consistency requirement. Separability requires that if two separate collections of selves  $U$  and  $U'$  each prefer the alternative  $a$  to the alternative  $b$ , then the combined collection of selves, obtained by merging collections of selves  $U'$  and  $U$ , also prefers  $a$  to  $b$ . Single-self respect and Separability together imply Pareto-optimality.

Continuity at indifferent selves requires strict preference orderings implied by the aggregator to be robust to the addition of nearly-indifferent collections of selves. This is the axiom that separates the class of aggregators we study from ordinal ones, since it implies that one self’s strict preference ordering is not reversed by adding an arbitrary (finite) number of selves, so long as the added selves are all close enough to being indifferent over the alternatives. This axiom only has meaning in a cardinal setting, and plays an analogous role in our setting as the Archimedean continuity axiom in expected utility theory. The axiom is weaker than requiring  $f$  (or the ordering of the alternatives implied by  $f$ ) to be continuous in the utilities of selves, for a fixed number of selves.

Finally, Duplication says that aggregation is only affected by the utility levels of the alternatives in a given choice set. In particular, choice is not affected by which of two alternatives is adjoined to a set as long as those two alternatives yield exactly the same utility to all of the selves. Note that it is not required that both of the elements added to the set  $A$  in P5 are from  $X \setminus A$ .

### 2.3 Examples of aggregators

The following are examples of context-dependent aggregators satisfying P1-P5, that are equivalent or closely related to models proposed in the existing literature.

**Example 1** (Reference Dependence). *Suppose that the aggregator is given by*

$$f(a, A, X, U) = \sum_{u \in U} (u(a) - \text{mean } u(A))^\rho,$$

where  $\rho$  is an odd integer and  $\text{mean } u(A)$  is a geometric or arithmetic mean over the set  $\{u(a')\}_{a' \in A}$ . This is a simple model of reference dependence.

**Example 2** (Passion-driven and passion-muted models). *Suppose there is a strictly monotonic and continuous weighting function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that for all  $U \in \mathcal{U}$  and choice sets  $A \subseteq X$ ,*

$$f(a, A, X, U) = \sum_{u \in U} g\left(\max_{b \in A} u(b) - \min_{b \in A} u(b)\right) u(a)$$

If  $g(\cdot)$  is increasing, the model is a passion-driven one in which selves who are more “passionate” about the alternatives in the set  $A$  receive greater weight in the decision-process because they are more vociferous than selves who are more or less indifferent among the possibilities. If  $g(\cdot)$  is decreasing, the model may be seen as a passion-muted model or a context-dependent version of the models of relative utilitarianism in Karni (1998), Dhillon and Mertens (1999), and Segal (2000), where a DM’s weight in society is normalized by her utility range over the grand set of alternatives.

Observe that  $a$  is preferred to  $b$  in the pair  $\{a, b\}$  if and only if

$$\sum_{u \in U} \underbrace{g(|u(a) - u(b)|)(u(a) - u(b))}_{\text{odd function of } u(a) - u(b)} > 0$$

Therefore, for pairwise choices the aggregator is similar to the additive difference model of Tversky (1969), which accounts for potentially intransitive pairwise choice behavior by positing utilities  $v_1, v_2, \dots, v_n$  and an odd  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  such that  $x \succ y$  if and only if  $\sum_{i=1}^n \phi(v_i(x_i) - v_i(y_i)) > 0$ . For larger choice sets, the aggregator can be thought of as a generalization of the additive difference model that permits context-dependence.

**Example 3** (Contextual concavity models from marketing). *Kivetz et al. (2004) (henceforth KNS) considers various models capturing the compromise effect documented in experimental settings. KNS consider goods (e.g., laptops) which have defined attribute levels (e.g., processor speed) and posit utility levels (“partworths”) for a given attribute. That is, they consider multiattribute alternatives and predefine the number of “selves” according to their selected good attributes. One type of model considered in KNS is referred to as a contextual concavity model. Using our notation, a symmetric version of the contextual concavity model they propose is given by*

$$f(a, A, X, U) = \sum_{u \in U} (u(a) - \min_{a' \in A} u(a'))^\rho,$$

where  $\rho$  is a concavity parameter.

## 2.4 Different types of selves

The examples of aggregators above all treat selves in the same way. However, many models in the existing literature propose methods of aggregation that treat some selves differently than others. For example, Fudenberg and Levine (2006) propose a dual-self impulse control model with a long-run self exerting costly self-control over a short-run self. One way to generalize this aggregator to any number of selves would be to introduce multiple types of short-term temptations, represented by selves  $u_1^{\text{sr}}, u_2^{\text{sr}}, \dots, u_n^{\text{sr}}$ , as well as a single long-run self  $u^{\text{lr}}$ .

Accommodating such type-dependent models of aggregation in our framework requires an extension of the framework and some extra notation, but no conceptual innovation. In particular, the definition of an aggregator must be extended to include a set of possible types, and the definition of a self must be extended to include a type. For ease of exposition, we restrict ourselves to the simplified framework in the main text and present the extension of the framework in [Appendix A](#). Our axioms and main theorem carry through to the extended framework.

## 2.5 Example: rationalizing the median procedure

The *median procedure* is a simple choice rule defined in Kalai et al. (2002). There is a strict ordering  $\succ$  defined over elements of  $X$ , and the DM always chooses the median element of each  $A \subseteq X$  according to  $\succ$  (choosing the right-hand side element among the medians from choice sets with even number of alternatives).

To rationalize this behavior, we consider the following aggregator.

$$f(a, A, X, U) = \prod_{u \in U} (u(a) + \max_{a' \in X} u(a') - \text{med}_{a' \in A} u(a')),$$

where  $\text{med}_{a' \in A} u(a')$  is the median element of the set  $\{u(a')\}_{a' \in A}$ , with the convention that in sets with an even number of distinct utility levels, the median is the smaller of the two median utility levels. The geometric aggregation implies that in case of selves having exactly the opposite preferences, the aggregated utility of an alternative from a given choice set is maximized when it is closest to the median element of the utility levels from the choice set.

Indeed, we claim that with the above aggregator, two selves can be used to rationalize the median procedure. Let  $a_1, a_2, \dots, a_N$  stand for the increasing ordering of alternatives in  $X$  according to  $\succ$ , and define  $u_1(a_i) = i + \varepsilon$  and  $u_2(a_i) = N + 1 - i$  for all  $i \in \{1, \dots, N\}$ . It is easy to see that for small enough  $\varepsilon > 0$  it is indeed one of the median elements of any choice set that maximizes  $f$ , since the sum of  $u_1(a) + \max_{a' \in X} u_1(a') - \text{med}_{a' \in A} u_1(a')$  and  $u_2(a) + \max_{a' \in X} u_2(a') - \text{med}_{a' \in A} u_2(a')$  is constant across all elements of  $X$ , and the aggregated utility is defined to be the product of the two terms.

This rationalization is relatively simple and intuitive: the above selves are defined such that the DM is torn between two motivations, one in line with ordering  $\succ$ , and one going in exactly the opposite direction. Moreover, the geometric aggregation of these preferences drives the DM to choose the most central element of any choice set. In contrast, Kalai et al. (2002) show that in their framework, in which exactly one self is responsible for any decision, as the size of  $X$  increases, the number of selves required to rationalize the median procedure goes to infinity.<sup>11</sup> While dictator-type aggregators do not provide an intuitively appealing explanation for the median procedure, an aggregator that captures compromise along selves along the lines of the above-defined  $f$  does yield a model that rationalizes the median procedure in a simple and intuitive way.

There are many variants of the above aggregator that given two selves with diametrically opposed interests do not select exactly the median from every choice set, but have a tendency to induce

---

<sup>11</sup>Another simple procedure they consider, which in their framework also requires a large number of selves to rationalize, is the second-best procedure, suggested first by Sen (1993). It is again possible to provide an aggregator fitting our framework such that two selves with opposite interests rationalize the choice rule - please contact the authors for details.

the choice of a centrally located element from any choice set. In general, if  $f$  is menu-dependent and aggregates the utilities of selves through a concave function, the choice induced by  $f$  exhibits a *compromise effect* or *extremeness aversion*, as in the experiments of Simonson (1989): given two opposing motivations, an alternative is more likely to be selected the more centrally it is located. If, on the other hand,  $f$  is menu-dependent and convex, then it can give rise to a *polarization effect*, as in the experiments of Simonson and Tversky (1992): the induced choice is likely to be in one of the extremes of the choice set. Hence, our model can be used to reinterpret experimental choice data in different contexts, in terms of properties of the aggregator function.

The aggregator and two selves in this example together rationalize a very specific type of behavior. However, the same aggregator might act differently on a different collection of selves. For example, if the two selves did not have exactly opposing preferences, the aggregator might not choose a centrally located alternative in every choice set. In the next sections we turn to studying the *set* of behaviors that an aggregator can rationalize (with different possible selves).

### 3 Counting IIA violations

The examples of decision-rules presented in the previous section violate the Independence of Irrelevant Alternatives (IIA) because they are context-dependent.<sup>12</sup> IIA requires that if  $a \in A \subset B$  and  $c(B) = a$  then  $c(A) = a$ . This says that if an alternative is chosen from a set, then it should be chosen from any subset in which it is contained. It is well known that a choice function can be rationalized as the maximization of a single preference relation if and only if it has no violations of IIA. In the next section we connect the set of choice functions that an aggregator can rationalize with  $n$  selves to the number of IIA violations that a choice function exhibits. To do this, we formally define an accounting procedure for the number of IIA violations.

The number of IIA violations can be determined straightforwardly for choice functions over three-element sets; e.g., if the choice over pairs is transitive but the second-best element according to the pairs is selected from the triple, there is one violation of IIA. For a larger set of alternatives, there are different plausible ways to define the number of violations. For example, suppose that

$$\begin{aligned} c(\{a, b, c, d, e, f\}) &= d \\ c(\{a, b, c, d, e\}) &= b \\ c(\{a, b, c, d\}) &= b \\ c(\{b, c, d\}) &= c. \end{aligned}$$

---

<sup>12</sup>Under the restriction of single-valued choice, the IIA condition is equivalent to Sen's  $\alpha$  - see Sen (1971) - or WARP, the weak axiom of revealed preference.

In light of  $c(\{a, b, c, d, e, f\}) = d$ , IIA dictates that the last three choices should be  $d$  (but they are not). In light of  $c(\{a, b, c, d, e\}) = b$ , IIA dictates that the choice from  $\{b, c, d\}$  should be  $b$  (but it is not), and the IIA implication for  $\{b, c, d\}$  is again violated in light of  $c(\{a, b, c, d\}) = b$ . Hence, one way of counting would indicate five IIA violations with respect to the above four choice sets.

However, according to our counting procedure, there are two IIA violations in this example: only the choices from  $\{a, b, c, d, e\}$  and  $\{b, c, d\}$  are associated with violations. The reason is that while  $c(\{a, b, c, d\}) = b$  does contradict  $c(\{a, b, c, d, e, f\}) = d$ , the intermediate choice  $c(\{a, b, c, d, e\}) = b$  itself implies by IIA that  $c(\{a, b, c, d\}) = b$ . With this motivation in mind, our accounting procedure associates an IIA violation with a choice set if and only if the choice from the set violates the choice from a superset that contains exactly one more alternative.

**Definition 2** (IIA violation). *The set  $A$  causes an IIA violation under the choice function  $c(\cdot)$  if (1) there exists  $B$  such that  $A \subset B$  and  $c(B) \in A \setminus \{c(A)\}$ , and (2) for every  $A'$  such that  $A \subset A' \subset B$ ,  $c(A') \notin A$ .*

Then, the total number of IIA violations is defined in the natural way.

**Definition 3** (Number of IIA violations). *The total number of IIA violations of a choice function  $c(\cdot)$  is given by  $\text{IIA}(c) = \#\{A \in P(X) \mid A \text{ causes an IIA violation}\}$ .*

The above definition can yield a large number of IIA violations for choice rules that can be defined relatively easily. Consider, for example, the choice function that arises when one strict preference ordering dictates choice whenever the menu contains some highlighted alternative, while an opposite strict preference ordering dictates choice in the absence of that alternative. In [Section 4.5](#) we provide a construction that collapses IIA violations compatible with each other into a single violation, and show how the construction can be used to sharpen our results.

There are other plausible measures for the number of IIA violations implied by a choice function. One alternative measure would be the minimal number of sets at which the choice function would have to be changed to make it rational. This measure can in general be either larger or smaller than our measure of the number of IIA violations.<sup>13</sup>

---

<sup>13</sup>Indeed, suppose that pairwise choices exhibit the transitive ranking  $a$  preferred to  $b$  preferred to  $c$ . Under our measure, there is one violation of IIA if  $c(\{a, b, c\}) = b$ , which is defeated once in the pair  $\{b, c\}$ , and two violations of IIA if  $c(\{a, b, c\}) = c$ , which is defeated twice. The alternative measure counts one violation either way. To see that the alternative measure can also be larger, consider the choice function over  $\{a, b, c, d, e\}$  which chooses the alphabetically-lowest alternative in all sets, except that  $b$  is chosen in three-element sets in which it is contained as well as from the pair  $\{a, b\}$ . The alternative measure counts four violations, while ours counts three. We thank both John Geanakoplos and Bart Lipman for suggesting this measure to us.

## 4 Main results

In this section we present our main results, which use our accounting procedure for IIA violations to characterize a lower bound on the set of choice functions that a model of multi-self aggregation can rationalize with  $n$  selves. For ease of exposition, in this section we restrict attention to aggregators where the aggregate utility of an alternative in a choice set  $A$  is independent of alternatives in  $X \setminus A$ .

**P6** (Independence of unavailable alternatives). *For any grand sets  $X, X' \in \mathcal{X}$  such that  $A \subseteq X \cap X'$ , and for any selves  $U^X \in \mathcal{U}(X)$  and  $U^{X'} \in \mathcal{U}(X')$  that agree on  $A$  (i.e.,  $U^{X'}(a) = U^X(a)$  for all  $a \in A$ ), the aggregator satisfies  $f(\cdot, A, X, U^X) = f(\cdot, A, X', U^{X'})$ .*

In [Supplementary Appendix C](#) we extend our results to a class of aggregators violating P6.

We start in [Section 4.1](#) by demonstrating how to construct selves that rationalize a choice function in the case of the passion-driven aggregator. The construction provides intuition for the connection between the number of selves and the number of IIA violations. In [Section 4.2](#) we generalize the construction to any aggregator satisfying a property that we call *triple-solvability*. This property holds for the aggregators in [Examples 1-3](#). In [Section 4.3](#), we provide sufficient conditions for triple-solvability within the class of additively separable and scale invariant aggregators. In particular, an aggregator is triple-solvable whenever it can rationalize two types of irrational behavior on a triple (choosing the worst pairwise element from the triple, and intransitive choice). We also argue that triple-solvability holds generically within this class of aggregators. In [Section 4.4](#), we point out that a weaker notion of triple solvability would suffice for our results. Finally, in [Section 4.5](#) we show that our construction can yield tighter bounds when the DM's IIA violations are related to each other in a systematic manner.

### 4.1 Rationalizing choice with passion-driven aggregation

Suppose that we are interested in rationalizing some choice function  $c(\cdot)$  using the passion-driven aggregator, which is given by

$$f(a, A, X, U) = \sum_{u \in U} g\left(\max_{b \in A} u(b) - \min_{b \in A} u(b)\right) u(a)$$

where  $g(\cdot)$  is increasing. Before considering an arbitrary grand set of alternatives  $X$ , let us first examine how this aggregator behaves on an arbitrary three-element set of alternatives  $\{a, b, c\}$ . In

particular, consider the collection of selves  $U = (u_1, u_2, u_3, u_4, u_5)$  specified below.<sup>14</sup>

$u_1$	$u_2$	$u_3$	$u_4$	$u_5$
$b \ 2$	$b \ 2$	$c \ 2$	$a, c \ 2$	$a \ 2$
$c \ 1$	$a \ 1$	$b \ 1$	$b \ 0$	$b, c \ 0$
$a \ 0$	$c \ 0$	$a \ 0$		

It is easy to verify that the aggregator selects  $a$  from the choice set  $\{a, b\}$ . Suppressing notational dependence on  $X = \{a, b, c\}$ , observe that  $f(a, \{a, b\}, U) = 4g(2) + g(1)$  and  $f(b, \{a, b\}, U) = 2g(2) + 3g(1)$ , hence  $f(a, \{a, b\}, U) > f(b, \{a, b\}, U)$  if and only if  $g(2) > g(1)$ , which holds since  $g(\cdot)$  is strictly increasing. By contrast, the aggregator assigns equal utility to all alternatives in any other menu:

$$\begin{aligned}
 f(a, \{a, c\}, U) &= f(c, \{a, c\}, U) = 2g(0) + g(1) + 2g(2) \\
 f(b, \{b, c\}, U) &= f(c, \{b, c\}, U) = 3g(1) + 2g(2) \\
 f(a, \{a, b, c\}, U) &= f(b, \{a, b, c\}, U) = f(c, \{a, b, c\}, U) = 5g(2)
 \end{aligned}$$

That is,  $a$  beats  $b$  when the choice set is  $\{a, b\}$ , while the selves cancel each other out for any other subset of  $\{a, b, c\}$ . We call such a collection of selves defined on  $\{a, b, c\}$  a *triple-basis* for this aggregator. In the case of this aggregator, the selves above would still be a triple-basis if we were to scale all the utilities by a common constant.

Given an arbitrary  $X$  and any choice function  $c$  defined on  $X$ , we can use the triple-basis above to construct a collection of selves that rationalize  $c$  using the passion-driven aggregator  $f$ . The procedure works as follows. We examine all possible choice sets in  $X$  from smallest to largest, first going through all choice sets of size two, then all choice sets of size three, etc. We ignore any choice set that does not cause an IIA violation. For each choice set  $A$  that does cause an IIA violation, the construction creates a collection of selves  $U^A$  defined on  $X$  such that

1.  $c(A)$  is selected under  $f \circ U^A$  from every subset of  $A$  in which it is contained.
2. The selves  $U^A$  cancel each other out under  $f$  on every other choice set (that is, on sets not containing  $c(A)$  or sets containing some element of  $X \setminus A$ ).
3. The selves  $U^A$  are “indifferent enough” so that their trickle-down effect does not overturn the strict preference of previously constructed selves.

Finally, the construction creates an extra self  $u^*$ , that is indifferent enough never to overturn any of the other selves’ strict preferences, in the standard way: the self allocates the highest utility

---

<sup>14</sup>In the  $i$ -th column, the alternative on the left is assigned the utility number to its right.

to  $c(X)$ , the next highest utility to  $X \setminus \{c(X)\}$ , and so on. All in all, this procedure constructs a collection of  $1 + 5 \cdot \text{IIA}(c)$  selves (one initial self, and five selves for each IIA violation).

Using the triple-basis above, it is easy to construct the collection of selves  $U^A$  associated with a set  $A$  that causes an IIA violation. To satisfy the first two properties above, we simply let  $c(A)$  play the role of  $a$  in the triple-basis, all the elements of  $A \setminus \{c(A)\}$  play the role of  $b$ , and all the elements of  $X \setminus A$  play the role of  $c$ . That is, we extend the utilities from  $\{a, b, c\}$  to the given  $X$  such that: each self allocates the same utility to  $c(A)$  as to  $a$  in the triple-basis, the same utility to elements of  $A \setminus c(A)$  as to  $b$  in the triple-basis, and the same utility to  $X \setminus A$  as to  $c$  in the triple-basis. Neutrality (P1) and duplication (P5) then imply that the properties of the triple-basis carry over: for each  $B \subseteq A$  that contains  $c(A)$ ,  $f(c(A), B, X, U^A) > f(y, B, X, U^A)$  for all  $y \in B$ , and for all other subsets  $B' \subseteq X$ ,  $f(x, B'^A) = f(y, B'^A)$  for all  $x, y \in B'$ . To satisfy the third property above, we can use continuity (P4) and scale all the selves in the triple-basis by some appropriately chosen  $\varepsilon > 0$ .

This entire collection of selves rationalizes  $c(\cdot)$  under  $f$ . The construction ensures that  $c(A)$  is selected from any set causing an IIA violation; one need only check that constructed selves do not interfere with choices associated with sets that do not cause IIA violations. To loosely illustrate the idea, consider any nested sequence of choice sets that decreases by one alternative. Given  $X$ , or any set which does not cause an IIA violation, all selves besides  $u^*$  are indifferent, hence by single-self respect (P2) and separability (P3) the preferences of  $u^*$  prevail. For the first set of the sequence that contradicts the choice from  $X$ , a triple-basis was created with selves passionate enough to overrule  $u^*$  and guarantee that the  $c$ -choice from this set is the  $f$ -maximizer (while the other triple-bases created will be indifferent). Similarly, whenever along the sequence there is a set that contradicts the choice of the previous set, another triple-basis was created that overrules the preferences of all selves created in association with larger sets.

The above construction implies that if we permit the model to have  $n$  selves, any choice function (on any grand set of alternatives) having fewer than  $\frac{n-1}{5}$  IIA violations can be rationalized using this aggregator.

## 4.2 Main rationalizability result

The construction from the previous subsection can be generalized to any aggregator having the property that there exists  $k \in \mathbb{Z}_+$  such that there exists a triple-basis consisting of  $k$  selves that are arbitrarily close to being indifferent. As we showed above, passion-driven aggregators satisfy this requirement with  $k = 5$ . This property is relatively simple to check for a concrete aggregator, since it is defined for a three-element set. For scale-invariant aggregators, which satisfy the property that measuring utilities in a different unit does not change the ordering implied by the aggregator,

checking the property is particularly simple, since it then suffices to construct one triple-basis which can be scaled as needed. See [Section 4.3](#) for an investigation of the prevalence of triple-solvability in the class of scale-invariant aggregators.

**Definition 4.** *A collection of selves  $\hat{U} \in \mathcal{U}(\{a, b, c\})$  is a triple-basis for  $f$  with respect to  $\{a, b, c\}$  if  $f(a, \{a, b\}, \{a, b, c\}, \hat{U}) > f(b, \{a, b\}, \{a, b, c\}, \hat{U})$ , and  $f(\cdot, A, \{a, b, c\}, \hat{U})$  is constant for all other  $A \subseteq \{a, b, c\}$ . Aggregator  $f$  is triple-solvable with  $k$  selves if there exists a triple  $\{a, b, c\}$  and  $k \in \mathbb{Z}_+$  such that for every  $\delta > 0$  there is  $U \in \mathcal{U}^k(\{a, b, c\})$   $\delta$ -indifferent with respect to  $\{a, b, c\}$  constituting a triple-basis for  $f$  with respect to  $\{a, b, c\}$ .*

Triple-solvability with  $k$  selves implies that we can find a sequence of triple-bases containing  $k$  selves that converge to an indifferent self. Our main theorem applies to all aggregators satisfying P1-P6 and triple-solvability.

**Theorem 1.** *Suppose  $f$  satisfies P1-P6 and is triple-solvable with  $k_f$  selves. Then, using  $n$  selves,  $f$  can rationalize any choice function  $c$ , defined on any finite grand set of alternatives  $X$ , that exhibits at most  $\frac{n-1}{k_f}$  IIA violations.*

The choice functions exhibiting no more than  $\frac{n-1}{k_f}$  IIA violations constitute a lower bound on behaviors that an aggregator  $f$  satisfying the conditions in the theorem can rationalize. The result can be restated such that if  $f$  satisfies P1-P6 and is triple-solvable with  $k_f$  selves then it can rationalize any choice function  $c$  with no more than  $1 + k_f \cdot \text{IIA}(c)$  selves. That is,  $1 + k_f \cdot \text{IIA}(c)$  is an upper bound on the number of selves (or complexity) required for rationalizing  $c$  with  $f$ .

One immediate implication of this theorem is that in spite of having a structured form, any aggregator satisfying the above properties can rationalize any choice function if sufficiently many selves are permitted by the model. In other words, if a model of multi-self decision-making *does not* restrict the number of selves and corresponds to an aggregator satisfying the properties above, then it generates a theory that cannot be refuted. The result therefore points out the importance of putting *a priori* restrictions on the number of selves in models of multi-self decision-making - a practice followed in some but not all of the existing literature.

### 4.3 Sufficient conditions for triple-solvability

Triple-solvability holds for all the aggregators featured in [Section 2.3](#).<sup>15</sup> The fact that these examples illustrate various models of multi-selves decision-making proposed in the literature suggests that the property holds broadly. In this section we verify this intuition within the class of aggregators that are additively separable and scale-invariant. These conditions imply that the aggregator can be written in the following form:

$$f(a, A, X, U) = \sum_{u \in U} f(a, A, X, u), \text{ where } f(a, A, X, \alpha u) = \phi(\alpha) f(a, A, X, u)$$

for some invertible and odd  $\phi : R \rightarrow R$ . Scale-invariance says the unit in which preference intensity is measured does not matter: the selves  $(\alpha u_1, \alpha u_2, \dots, \alpha u_n)$  are aggregated in the same way as the selves  $(u_1, u_2, \dots, u_n)$ . Additive separability is a common functional form assumption.<sup>16</sup> This class contains, for instance, the reference-dependent aggregator highlighted in [Example 1](#).

We establish two formal theorems, which formalize in different ways the statement that triple-solvability is prevalent in this class of aggregators.

The first result shows that if an aggregator can rationalize two types of simple irrational choice on a grand set of alternatives consisting of three elements, then it is triple-solvable, and can therefore rationalize any choice function on any grand set of alternatives.

**Theorem 2.** *Suppose  $f$  satisfies P1-P6, it is additively separable and scale-invariant, and it can rationalize the following two behaviors on some triple of alternatives  $\{a, b, c\}$ :*

1. *Intransitive choice; and*
2. *Transitive choice, where the worst pairwise element is best in the triple.*

*Then there is  $k_f$  such that  $f$  is triple-solvable with  $k_f$  selves.*

The second result shows that an aggregator from the above class is triple-solvable with five selves unless one of two particular linear relationships hold for aggregated utilities of alternatives given different choice sets, for *every* self defined over a triple.

<sup>15</sup>Solvability of the reference-dependent aggregator in [Example 1](#) will be shown by the results in this section. For the contextual concavity aggregator in [Example 3](#), the following constitutes a triple basis for any  $\rho \neq 1$ :

$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$
$a$ 4	$a$ 3	$b$ 4	$b, c$ 3	$c$ 3	$c$ 4
$b$ 3	$c$ 2	$a$ 3	$a$ 1	$a$ 2	$b$ 2
$c$ 1	$b$ 1	$c$ 1		$b$ 1	$a$ 1

<sup>16</sup>For properties implying additive separability, see Debreu (1959, Theorem 3) and Maskin (1978).

**Definition 5.** An aggregator  $f$  is degenerate if for every single self  $u$  over a triple  $\{a, b, c\}$ ,  $f \circ u$  satisfies at least one of the following two linear equalities (suppressing notational dependence on  $X = \{a, b, c\}$ ):

1.  $f$  does not stretch utility differences over pairs:

$$\begin{aligned} & f(a, \{a, c\}, u) - f(c, \{a, c\}, u) \\ &= f(a, \{a, b\}, u) - f(b, \{a, b\}, u) + f(b, \{b, c\}, u) - f(c, \{b, c\}, u); \end{aligned}$$

2. How  $f$  ranks alternatives in the triple is determined by the pairwise rankings:

$$\begin{aligned} & [f(a, \{a, b, c\}, u) - f(b, \{a, b, c\}, u)] + [f(a, \{a, b, c\}, u) - f(c, \{a, b, c\}, u)] \\ &= [f(a, \{a, b\}, u) - f(b, \{a, b\}, u)] + [f(a, \{a, c\}, u) - f(c, \{a, c\}, u)]. \end{aligned}$$

The aggregator is non-degenerate if neither equation holds for some self.

The first property is a form of degenerate menu dependence; knowing how the aggregator acts on any two pairs, one may immediately recover its behavior on the remaining pair. The second property is a different form of degenerate menu dependence, suggesting that preference intensity over two alternatives is independent of the size of the choice set. The additive aggregator defined by  $f(a, A, X, u) = \alpha u(a) + \beta_A$ , where  $\alpha > 0$  and  $\beta_A$  are constants, is degenerate and effectively reduces to simple utilitarianism.

It is straightforward to check whether an aggregator is non-degenerate: one need only exhibit a single self defined over a triple  $\{a, b, c\}$  for which neither of the above linear conditions holds. For example, using the single self  $u$  given by  $u(b) = 4 > u(c) = 2 > u(a) = 1$  shows that the version of [Example 1](#) with an arithmetic mean is non-degenerate. Then the theorem below implies that the aggregator must be triple solvable.

**Theorem 3.** *If an additive and scale-invariant aggregator satisfying P1-P6 is non-degenerate then it is triple-solvable with five selves.*

In [Supplementary Appendix A](#) we define a natural topology over scale-invariant and additive aggregators, and show that the set of aggregators that are non-degenerate is open and dense. Hence, non-degeneracy — and therefore triple-solvability with at most five selves — is a generic property of these aggregators.

#### 4.4 A weaker notion of solvability

While triple solvability is a property that is broadly satisfied, it can be seen from our construction that our theorem would still hold under a weaker condition. It suffices that there exist a collection of selves that are arbitrarily close to being different on all but some set  $\{a, b\}$ . We formalize this idea in [Supplementary Appendix B](#), where we extend the notion of a triple-basis to an *approximate triple-basis*. For some aggregators, approximate triple-solvability can yield a triple-basis with a drastically smaller number of selves. Indeed, consider an aggregator of the form

$$f(a, A, X, U) = \sum_{u \in U} h(\max_{a' \in A} u(a'))u(a),$$

where  $\lim_{x \rightarrow \infty} h(x)x = 0$ . Under such an aggregator, the presence of an alternative with very high utility for a self means that self is given less say in the decision process (a “populist”-type model). This can be used to create a *single-self* approximate triple-basis  $u$ : let  $u(a)$  and  $u(b)$  such that  $f(a, \{a, b\}, \{a, b, c\}, u) - f(b, \{a, b\}, \{a, b, c\}, u) = \delta$  (for small enough  $\delta$  this is always possible), and let  $u(c)$  be high enough so that  $u$  is  $\varepsilon$ -indifferent between any two elements given sets containing  $c$ . [Theorem 2](#) then implies that using  $n$  selves, the aggregator can rationalize all choice functions with no more than  $n - 1$  IIA violations.

#### 4.5 Systematic IIA violations

Our construction allocates a different triple-basis (or approximate triple-basis) for every IIA violation. However, there can be IIA violations that are “in the same direction” (that do not contradict each other). In this case, parts of the associated triple-bases in our construction can be combined (or *collapsed*) together to yield tighter bounds.

For example, recall the triple-basis for the passion-driven aggregator, and fix some alternative  $a$ . Every time the choice of  $a$  from some set causes an IIA violation, the triple-basis constructed has a self  $u_5$  in which  $a$  is preferred to  $X \setminus \{a\}$ , all elements of which are indifferent to each other. Under the passion-driven aggregator, all of the  $u_5$  selves constructed when the choice was  $a$  can be collapsed into a single-self. More generally, the following corollary to [Theorem 1](#) holds.

**Corollary 4.** *Suppose  $f$  satisfies P1-P6 and is triple-solvable with  $k_f$  selves. For an arbitrary choice function  $c$ , defined on any finite grand set of alternatives  $X$ , let*

$$D(c) = \#\{a \in X \mid c(A) = a \text{ for some } A \subseteq X \text{ causing an IIA violation}\}$$

*be the number of distinct elements whose choice is associated with an IIA violation. Then, the number of selves needed to rationalize the choice function  $c$  is at most  $1 + \ell \cdot D(c) + m \cdot \text{IIA}(c)$ ,*

where  $\ell + m \leq k_f$ .

This effect is particularly pronounced when the triple-basis has only one self, as in the approximately triple-solvable aggregators introduced above. To illustrate this, consider the following example: let  $x^* \in X$ , and let  $\succ_1$  and  $\succ_2$  be strict orderings on  $X$  such that  $x \succ_1 x^*$  and  $x \succ_2 x^*$  for every  $x \in X \setminus \{x^*\}$ , and  $y \succ_1 x$  for  $x, y \in X \setminus \{x^*\}$  if and only if  $x \succ_2 y$ . Consider a decision-maker who from choice sets not containing  $x^*$  selects the best element according to  $\succ_1$ , but from choice sets containing  $x^*$  selects the best element according to  $\succ_2$ . This behavior describes, for example, a customer in a restaurant who chooses the tastiest item from a menu if the menu does not contain onion rings, while choosing the healthiest item in the presence of onion rings, because they are so greasy as to make the customer feel guilty about his eating habits.<sup>17</sup> The above simple behavior generates a large number of IIA violations if  $X$  is large.<sup>18</sup> However, these IIA violations do not contradict each other: if choice from set  $B$  contradicts the choice from  $A \supset B$ , then there is no  $B' \subset B$  such that the choice from  $B'$  contradicts the choice from  $B$ . As we show below, this can be used to merge all collections of selves into a single collection, drastically reducing the number of selves required to rationalize the above choice function.

Consider the aggregator introduced in the previous subsection, which was shown to be approximately triple-solvable with a single self. Our construction calls for (i) creating a self whose utility function is in line with  $\succ_2$ ; and (ii) creating a self for all sets associated with an IIA violation, such that the self attaches high enough utility to  $x^*$  such that the self becomes close enough to indifferent in the presence of  $x^*$ , and among the other alternatives allocates the highest utility to the choice from the given set. The latter selves can all be collapsed into a single self, such that the utility function of the self is in line with  $\succ_1$  over  $X \setminus \{x^*\}$  (while keeping the utility of  $x^*$  at a level that makes the self nearly indifferent in the presence of  $x^*$ ). Our construction then implies that the above choice function can be rationalized with two selves. This is clearly a tight bound.

## 5 The meaning of mistakes in a Strotzian model

In a seminal paper, Strotz (1955) models a DM who acts in anticipation of the choice of a future self. Gul and Pesendorfer (2001) contains a time-consistent interpretation and axiomatization of the Strotzian model, where a DM commits to a menu in anticipation of having to choose from that menu subject to temptation. In the language of this paper, such a DM selects a menu  $A$  that maximizes  $W(c(A))$ , where  $W$  is a utility function over the alternatives and  $c$  is the *rational* choice

---

<sup>17</sup>We thank Ran Spiegler for suggesting this choice rule.

<sup>18</sup>The number of IIA violations is  $2^{n-1} - n - 1$ : the choice from every set  $B$  having at least two elements and not containing  $x^*$  contradicts the choice from  $B \cup \{x^*\}$ .

function that corresponds to choosing the most tempting alternative;  $c$  is rational because a DM in their framework has only a *single* temptation ranking.

In this section we propose a *generalized* Strotzian model accommodating the possibility that  $c$  is not a rational choice function, and study its properties using the results of Section 4.<sup>19</sup> Denoting the grand set of alternatives by  $X$ , the DM has a preference relation  $\succeq$  over menus (nonempty elements of  $P(X)$ ). When evaluating a menu, the DM takes into account that her choice from that set will be governed by *multiple*, possibly conflicting interests. Consider the following utility representation capturing this behavior.

**Definition 6.** *The DM’s preference over menus  $\succeq$  has a generalized Strotzian representation if there exists a collection of selves  $U \in \mathcal{U}(X)$ , an aggregator  $f$ , and a utility function  $W : X \rightarrow \mathbb{R}$  on alternatives such that  $\succeq$  is represented by the utility function  $V : P(X) \rightarrow \mathbb{R}$  on sets, defined by*

$$V(A) = W\left(\arg \max_{a \in A} f(a, A, X, U)\right).$$

The generalized Strotzian representation has a straightforward interpretation: the DM simply picks the set from which the element foreseen to be chosen yields the greatest current utility. However, the DM chooses the best menu subject not to the choice of one motivation, but rather the choice maximizing an aggregate of multiple motivations.<sup>20</sup>

The following three axioms characterize a DM with a generalized Strotzian representation.

**Axiom 1** (Preference Relation)  $\succeq$  is complete and transitive.

**Axiom 2** (Strict Ordering)  $\succeq$  is a strict ordering on the singletons  $\{a\}_{a \in X}$ .

**Axiom 3** (IUUA) For all  $A \in P(X)$ , there exists  $a \in A$  such that  $A \sim \{a\}$ .

In the classical theory of choice, a set is assumed to be indifferent to its best element. The IUUA axiom — short for *Independence of Utility to Unchosen Alternatives* — retains the idea that the set is indifferent to the “best” element inside it, even if that element may not arise from a menu-independent ranking. That is, IUUA permits context-dependent behavior without introducing psychological costs (e.g., through temptation, as in Gul and Pesendorfer (2001)).

**Theorem 5.**  *$\succeq$  satisfies Axioms 1-3 if and only if  $\succeq$  has a generalized Strotzian representation using a collection of selves  $U$  and an aggregator  $f$  satisfying P1-P6 and triple-solvability with  $k$  selves. Moreover, defining the DM’s “anticipated” choice function  $c_{\succeq}$  (induced by  $\succeq$ ) by*

$$c_{\succeq}(A) = a \text{ if } a \in A \text{ and } A \sim \{a\},$$

---

<sup>19</sup>We thank Eddie Dekel for suggesting the Strotzian interpretation of the representation.

<sup>20</sup>The above conception bears a relation to the separation of decision utility and experienced utility proposed by Kahneman, Wakker and Sarin (1997).

the number of selves in the representation is no larger than  $1 + k \cdot \text{IIA}(c_{\succeq})$ .<sup>21</sup>

*Proof.* To prove this theorem, note that Axioms 1-3 together ensure that we may uniquely define  $c_{\succeq}$  as above. Because each menu is indifferent to the alternative chosen by the induced choice function, the DM’s preferences over menus may be represented by a utility function  $W(\cdot)$  over the alternatives in  $X$ . We may then use the result of [Theorem 1](#) to rationalize the induced choice function. ■

The bound on the number of selves using the number of anticipated IIA violations raises connections to the literature on choice over menus. The generalized Strotzian model implies that for any pair  $\{a, b\}$ , either  $\{a, b\} \sim \{a\}$  or  $\{a, b\} \sim \{b\}$ . However, for larger sets, it may be that  $A \cup B \succ A, B$  (behavior which is interpreted as a preference for flexibility in Kreps (1979)), that  $A, B \succ A \cup B$ , or that  $A \succ A \cup B \succ B$  (as in Gul and Pesendorfer (2001)’s *Betweenness*, which they interpret in terms of costly self-control). The interpretation here is different from all of the above:

**Observation 1.** *If  $A \cup B$  is not indifferent to either  $A$  or  $B$  then an IIA violation necessarily occurs in the anticipated choice function  $c_{\succeq}$ .*

A generalized Strotzian DM is conflicted when she makes her choice from the menu, and depending on how she resolves the compromise among selves, might prefer a larger or smaller set that leads to a better choice according to the utility  $W$ . How  $A \cup B$  stands in relation to  $A$  and  $B$  provides information as to when the DM expects to be conflicted; and when an IIA violation occurs, the upper bound on the minimal number of selves required to rationalize the behavior using a triple-solvable aggregator increases. Although the generalized Strotzian representation is not contained within the class of utilities considered by Dekel, Lipman and Rustichini (2001), this observation is related to their result that the subjective state space in a model of unforeseen contingencies grows when there is additional desire for flexibility or self-control. IIA violations in anticipated choice are precisely ruled out in Gul and Pesendorfer (2001) by their *No Compromise* axiom, which is more restrictive than Axiom 3 because it requires that  $A \cup B \sim A$  or  $A \cup B \sim B$  for all menus  $A, B$  – thereby leading to a single temptation ranking. By contrast, in our setting “anticipated” IIA violations reveal additional conflicting motivations.

---

<sup>21</sup>Whether the “anticipated” choice is the actual choice made is an issue of consistency by the DM and her ability to foresee her future motivations; this can only be tested by observing actual choice *from* the menu. The interpretation of anticipation is not necessary for the model but the representation is suggestive of it.

## 6 Microeconomic models of household choice

Empirical evidence on household demand strongly suggests that it cannot arise from the maximization of a single utility. An extensive literature examines the microeconomic implications of collective choice in households where each member is a utility maximizer; and in particular, a branch of this literature examines such models under the restriction of Pareto-efficient household behavior. One question addressed in this setting is, given a household demand function over  $N$  goods, when do there exist  $n$  utility functions  $\{u_i\}_{i=1}^n$  and a continuously differentiable function  $\mu$  of prices and income such that the demand arises from the weighted utilitarian maximization of  $\sum_{u \in U} \mu(\text{price}, \text{income}) u(\cdot)$  given the budget set (i.e., weights and preferences vary independently). Browning and Chiappori (1998) show that if there are  $N$  goods, then any demand data can be explained by an  $(N - 1)$ -person household. In addition, to explain a given demand function using  $n$  people, it is necessary and sufficient that the rank of a certain matrix in a pseudo-Slutsky matrix decomposition be  $n - 1$ , though without further restrictions there can be a continuum of explanatory  $n$ -person models (Chiappori and Ekeland (2006)).<sup>22</sup>

To apply our framework in this context, we reinterpret selves as individuals of the household, and the aggregator as the mechanism that translates the individuals' preferences to household choice (this might be the outcome of a particular household bargaining procedure). Our approach differs in a number of ways from Browning and Chiappori (1998) and Chiappori and Ekeland (2006). First, the aggregator need not be weighted utilitarianism. Second, we address the question of rationalization by a concrete aggregator, while the above papers assume that the modeler does not know the underlying aggregation rule of the household, only that it belongs to the class of weighted utilitarian aggregators. Finally, we examine choice functions instead of demand functions. However, given that demand data is typically finite, suppose we denote by  $X$  the (finite) set of all available allocations, let each budget set correspond to a subset  $A \subset X$ , and identify the demand data with a function  $c$  that selects the allocation  $c(A)$  in the budget set  $A$ . Then, rationalizing the demand data corresponds to rationalizing an *incomplete choice function*:  $c$  renders a choice to any subset  $A$  of  $X$  for some collection of subsets  $\mathcal{A} \subset 2^X$ , but data on choices from sets in  $2^X \setminus \mathcal{A}$  is missing. As we show below, our results can easily be extended to arbitrary incomplete choice functions.

Rationalizing an incomplete choice function  $c$  with aggregator  $f$  implies finding a collection of selves  $U \in \mathcal{U}(X)$  such that  $f(c(A), A, X, U) > f(a, A, X, U)$  for all  $a \in A \setminus \{c(A)\}$  and  $A \in \mathcal{A}$  (it does not matter what choices  $f$  and  $U$  imply from sets in  $2^X \setminus \mathcal{A}$ ). To see how our theorems generalize, observe that the only element of the construction that needs to be modified is the

---

<sup>22</sup>The pseudo-Slutsky matrix is formally defined in Chiappori and Ekeland (2006); the rank condition they give,  $SR(n - 1)$ , is that this matrix can be decomposed as the sum of a symmetric negative semi-definite matrix and another matrix of rank at most  $n - 1$ . One intuition for the proof, which relies on exterior differential calculus, is that the Pareto-frontier for  $n$  people is  $n - 1$  dimensional, and weights and preferences can be varied independently.

number of IIA violations: in this more general context we say that an IIA violation is associated with choice set  $A \in \mathcal{A}$  if there is a nested sequence of choice sets  $A_1, A_2, \dots, A_k$  such that  $A_1 = X$ ,  $|A_j| - |A_{j+1}| = 1 \forall j \in \{1, \dots, k-1\}$ , and  $A_k = A$  for which the choice from  $A_k$  contradicts the choice from  $A_l$  for some  $l < k$ , and  $A_{l'} \notin \mathcal{A}$  for any  $l < l' < k$ . It is easy to see that this definition reduces to the original one in case of no missing data. Once the definition of  $\text{IIA}(c)$  is modified accordingly, it can be shown that [Theorem 1](#) holds (the proof is analogous).<sup>23</sup>

This means that for any aggregator satisfying our conditions, the demand data can be rationalized if there are sufficiently many people in the household. This complements the result obtained in Browning and Chiappori (1998) and Chiappori and Ekeland (2006), in that even if the researcher knows how preferences in the household are aggregated, if the number of individuals in the (extended) household is large or unknown, then the model does not imply any testable restrictions on household demand. Our combinatorial approach also permits a simple lower bound on demand data that a household with a known number of individuals can generate, in terms of the number of IIA violations implied by the demand data.

## 7 Discussion

The framework we propose in this paper provides a flexible environment for axiomatic investigation of multi-self models. As we pointed out, many of the models proposed in the existing literature can be translated into our framework such that the resulting aggregators satisfy the basic axioms we posited. However, there are other classes of aggregators that might be of interest, for example ordinal ones, which do not satisfy all our axioms. Our framework can still be useful to examine these aggregators, only some of our axioms need to be replaced by axioms that reflect the defining characteristics of the aggregators at hand. Furthermore, our set of axioms can also be supplemented with additional ones, leading to more specific classes of aggregators instead of the broad class of aggregation rules that we investigated in this paper, and hence to sharper predictions on implied choice with a fixed number of selves. We leave this direction, as well as extending our framework to dynamic settings, to future research.

---

<sup>23</sup>We note that  $\text{IIA}(c)$  for an incomplete choice function might be strictly less than  $\text{IIA}(\hat{c})$  for any completion  $\hat{c}$  of  $c$ . That is, it can be that any way of specifying choices for sets in  $2^X \setminus \mathcal{A}$  creates new IIA violations. Nevertheless, our theorems apply.

# Appendices

## A Non-anonymous aggregators

We extend our framework to incorporate aggregators that treat different selves in a non-anonymous manner, and show how our main result extends to this more general class of aggregators. The description of a self is extended by an abstract type, and the definition of an aggregator is extended to include a set of possible types. The abstract set of types could include, for example, “long-run” and “short-run” selves, or selves caring about different types of objectives, such as the “parental” and “work” selves mentioned in [Section 1](#).

An aggregator  $F = (T, f)$  specifies a set of possible types  $T$  and a function  $f$  that specifies the aggregate utility for every alternative  $a$  in every choice set  $A$ , given any (finite) grand set of alternatives  $X$  and any collection of selves  $S$  defined over  $X$  and  $T$ . A single self  $s$  is given by a pair  $(u, t)$ . For each positive integer  $n$ , we denote by  $\mathcal{S}^n(X, T)$  the set of all collections of selves (unordered lists) defined with respect to  $X$  and  $T$ , and let  $\mathcal{S}(X, T) = \cup_{n=1}^{\infty} \mathcal{S}^n(X, T)$ . We will denote a particular collection of selves by  $S$ , and refer to the selves in the collection as  $s_1, \dots, s_n$ . To denote the number of selves in  $S$ , we use the notation  $|S|$  or simply  $n$  when no confusion would arise.

This extension allows us to consider asymmetric aggregators.

**Example 4** (Asymmetric contextual concavity model). *Interpret each self as corresponding to a product attribute, for which the preference belongs to a certain type. The class of preferences is parametrized by a concavity index. The contextual concavity aggregator in Kivetz et al. (2004) is given by*

$$f(a, A, X, S) = \sum_{s \in S} (u(a) - \min_{a' \in A} u(a'))^{\rho(t)},$$

where  $\rho : T \rightarrow \mathbb{R}$  gives the concavity parameter for a type- $t$  self.

Since collections of selves are still defined as unordered lists, by construction aggregators in this framework treat selves of the same type symmetrically. Hence, asymmetries can enter only through different specified types. In particular, the framework constructed in the main text can be viewed as a special case of the extended framework proposed above, when the set of possible types is a singleton. Axioms P1-P6 can be generalized in a straightforward manner to the extended setting. Since the only changes required in the generalization are notational (all statements applying previously to selves now apply to the extended notion of a self), we omit restating the axioms in the extended framework. The main theorem is unchanged. The definition of a triple-basis is unchanged, as is the theorem:

**Theorem 6.** *Suppose  $f$  satisfies P1-P6 and is triple-solvable with  $k_f$  selves. Then, using  $n$  selves,  $f$  can rationalize any choice function  $c$ , defined on any finite grand set of alternatives  $X$ , that exhibits at most  $\frac{n-1}{k_f}$  IIA violations.*

Consider a different type of example.

**Example 5** (Costly self-control aggregators). *Fudenberg and Levine (2006) propose a dual-self impulse control model with a long-run self exerting costly self-control over a short-run self. The reduced-form model they derive has an analogous representation in our framework, with two selves: the long-run self, with utility given by  $u^{\text{lr}}$  (the expected present value of the utility stream induced by the choice in the present), and the short-run self, with utility function  $u^{\text{sr}}$  (the present period consumption utility).<sup>24</sup> Using our terminology, the reduced form representation of their model assigns to alternative  $a$  the aggregate utility  $u^{\text{lr}}(a) - C(a)$ , where term  $C(a)$  depends on the attainable utility levels for the short-run self and is labeled as the cost of self-control. For example, using Fudenberg and Levine (2006)'s parametrization,  $C(a) = \gamma[\max_{a' \in A} u^{\text{sr}}(a') - u^{\text{sr}}(a)]^\psi$ .*

*One way to generalize this aggregator to any number of selves would be to introduce multiple short-term temptations, represented by selves  $u_1^{\text{sr}}, \dots, u_n^{\text{sr}}$ , and to define the aggregator*

$$f(a, A, X, S) = u^{\text{lr}}(a) - \sum_{s \in S} \gamma[\max_{a' \in A} u^{\text{sr}}(a') - u^{\text{sr}}(a)]^\psi.$$

*Here, the long-run self is treated differently than the rest.*

As in the above generalization of Fudenberg and Levine (2006), it may be the case that a multi-self model places restrictions on how many selves of each type can appear. If types are restricted, the description of the model should also include a set of possible collections of types  $\mathcal{C}$ , given by a subset of the set of all possible unordered  $n$ -long lists of elements of  $T$ , for every  $n \in Z_+$ . The aggregator  $f$  need only specify the aggregate utility arising for any collection of selves  $S$  defined over  $X$  and  $T$  for which the implied collection of types is in  $\mathcal{C}$ .

Our results can be extended in a variety of ways to accommodate such restrictions. The most straightforward one imposes an assumption on the set  $\mathcal{C}$  (which is satisfied in [Example 5](#)). Assume the existence of a type  $t$  and a collection of types  $\widehat{T}$  such that appending any number of  $t$ -types to  $\widehat{T}$  results in a collection of types in  $\mathcal{C}$ . In the generalized costly self-control aggregator above, the short-run type being  $t$  and the singleton set of a long-run type as  $\widehat{T}$  satisfy this requirement.

---

<sup>24</sup>The long-run self's utility is equal to the short-run self's utility plus the expected continuation value induced by the choice. If the latter can take any value, then  $u^{\text{lr}}$  is not restricted by the short-run utility  $u^{\text{sr}}$ . If continuation values cannot be arbitrary (for example they have to be nonnegative) then  $u^{\text{sr}}$  restricts the possible values of  $u^{\text{lr}}$ , hence  $U$  has a restricted domain. In Fudenberg and Levine (2006) the utility functions also depend on a state variable  $y$ . Here we suppress this variable, instead make the choice set explicit.

Let  $T_t^n$  denote the collection of  $n$   $t$ -types. An aggregator  $f$  is *expandable* with  $t \in T$  from  $\widehat{T} \in \mathcal{C}$  if  $(\widehat{T}, T_t^n) \in \mathcal{C}$  for every  $n \in \mathbb{Z}_+$ . For an aggregator that is expandable with  $t$  from  $\widehat{T}$  we can define triple-solvability with  $k$  type- $t$  selves from  $\widehat{T}$  as the existence of a collection of selves consisting of  $|\widehat{T}|$  exactly indifferent selves over the triple whose type-composition is as in  $\widehat{T}$  and  $k$   $\delta$ -indifferent selves of type  $t$ , such that the above collection of types constitutes a triple-basis for every  $\delta > 0$ .

Given the above definitions, the following result is obtained.

**Theorem 7.** *Suppose  $f$  is triple-solvable with  $k$  type- $t$  selves from  $\widehat{T}$ . Then, using  $n$  selves,  $f$  can rationalize any choice function  $c$ , defined on any finite grand set of alternatives  $X$ , that exhibits at most  $\frac{n-1-|\widehat{T}|}{k}$  IIA violations.*

Because the aggregation term for a short-run self is the negative of the symmetric contextual concavity aggregation, it is immediate that the generalized costly self-control aggregator defined above is triple-solvable according to the extended definition.

## B Proofs

### Proof of Theorem 1

For an arbitrary choice function  $c$  we will construct a collection of  $1 + k \cdot \text{IIA}(c)$  selves which will be shown to rationalize  $c$ . This implies the claim in the theorem. In particular, we will construct  $k$  selves for each set with which an IIA violation is associated, and an extra self for  $X$ .

Let  $I_1 = \{A_1^1, \dots, A_{i_1}^1\}$  be the subsets of  $X$  such that there is an IIA violation associated with the set, but there is no proper subset of the set with which an IIA violation is associated. For  $j \geq 2$ , let  $I_j = \{A_1^j, \dots, A_{i_{j+1}}^j\}$  be the subsets of  $X$  such that there is an IIA violation associated with the set, but there is no proper subset of the set outside  $\bigcup_{l=1}^{j-1} I_l$  with which an IIA violation is associated. Let  $j^*$  be the largest  $j$  such that  $I_j \neq \emptyset$ .

We will now iteratively construct a collection of  $k$  selves for each set associated with an IIA violation, starting with sets in  $I_1$ . Consider any collection of  $k$  selves  $\bar{U}^1 = (\bar{u}_1^1, \dots, \bar{u}_k^1)$  that solves the triple  $\{a, b, c\}$  (the existence of such a triple follows from triple-solvability). For every  $A \subset I_1$ , construct now the following collection of selves  $U^A = (u_1^A, \dots, u_k^A)$ :

$$u_i^A(x) = \begin{cases} \bar{u}_i^1(a) & \text{if } x = c(A) \\ \bar{u}_i^1(b) & \text{if } x \in A, x \neq c(A) \\ \bar{u}_i^1(c) & \text{if } x \notin A \end{cases}$$

for every  $i = 1, \dots, k$ .

Suppose now that  $U^A$  is defined for every  $A \in \bigcup_{k=1}^j I_k$  for some  $j \geq 1$ . Let  $U_k$  be the collection of selves  $U_k = (U^{A_1^k}, \dots, U^{A_{i_k}^k})$ , for  $k = 1, \dots, j$ . Let  $\widehat{U}_j = (U_1, \dots, U_j)$ . By P4, there exists  $\delta > 0$  such that for any  $\delta$ -indifferent collection of  $k$  selves  $U'$ ,

$$f(a, A, X, \widehat{U}_j) > f(b, A, X, \widehat{U}_j) \text{ implies } f(a, A, X, (\widehat{U}_j, U')) > f(b, A, X, (\widehat{U}_j, U')).$$

Then by P3 and P6, we know

$$\begin{aligned} f(a, A, X, \widehat{U}_j, \widetilde{U}_1, \dots, \widetilde{U}_m) > f(b, A, X, \widehat{U}_j, \widetilde{U}_1, \dots, \widetilde{U}_m) \text{ implies} \\ f(a, A, X, (\widehat{U}_j, \widetilde{U}_1, \dots, \widetilde{U}_m, U')) > f(b, A, X, (\widehat{U}_j, \widetilde{U}_1, \dots, \widetilde{U}_m, U')) \end{aligned}$$

for any  $\widetilde{U}_1, \dots, \widetilde{U}_m$  collections of (exactly) indifferent selves.

Let now  $I_{j+1} = \{A_1^1, \dots, A_{i_{j+1}}^1\}$  be the subsets of  $X$  such that there is an IIA violation associated with the set, but there is no proper subset of the set outside  $I_j$  with which an IIA violation is associated. By triple-solvability with  $k$  selves, there is a  $\delta$ -indifferent collection of  $k$  of selves  $\bar{U}^{j+1} = (\bar{u}_1^{j+1}, \dots, \bar{u}_k^{j+1})$  that solves the triple  $\{a, b, c\}$ . For every  $A \subset I_{j+1}$ , construct now the following collection of selves  $U^A = (u_1^A, \dots, u_k^A)$ :

$$u_i^A(x) = \begin{cases} \bar{u}_i^{j+1}(a) & \text{if } x = c(A) \\ \bar{u}_i^{j+1}(b) & \text{if } x \in A, x \neq c(A) \\ \bar{u}_i^{j+1}(c) & \text{if } x \notin A \end{cases}$$

for every  $i = 1, \dots, k$ . Let  $U_{j+1}$  be the collection of selves  $(U_j, U^{A_1^1}, \dots, U^{A_{i_{j+1}}^1})$ .

The above procedure generates a collection of  $k \cdot \text{IIA}(c)$  selves in  $j^*$  steps. Then by P3 and P4 there is  $\delta_{j^*} > 0$  such that for any  $\delta_{j^*}$ -indifferent  $u$ ,  $f(a, A, X, U_{j^*}) > f(b, A, X, U_{j^*})$  implies  $f(a, A, X, (U_{j^*}, u)) > f(b, A, X, (U_{j^*}, u))$ . Finally, construct one more self the following way: let  $a_1 = c(X)$  and  $a_k = c(X \setminus \{a_1, a_2, \dots, a_{k-1}\})$  for  $2 \leq k \leq n$ . Construct  $u^* : X \rightarrow \mathbb{R}$  such that  $u^*(a_1) > u^*(a_2) > \dots > u^*(a_n)$  and  $u^*$  is  $\delta_{j^*}$ -indifferent.

We show the collection of selves  $U_c \equiv (U_{j^*}, u^*)$  rationalize  $c$  with aggregator  $f$ .

**Observation 2.** *For any set  $A$  with which there is an IIA violation associated, by the construction of  $U^A$  and by P1 and P5,  $f(a, B, X, U^A) = f(b, B, X, U^A) \forall a, b \in B$  and  $B$  such that either  $B \setminus A \neq \emptyset$  or  $c(A) \notin B$ , and  $f(c(A), B, X, U^A) > f(b, B, X, U^A) = f(b', B, X, U^A) \forall b, b' \in B \setminus \{c(A)\}$  and  $B$  such that  $B \setminus A = \emptyset$  and  $c(A) \in B$ .*

We will now show that the choice induced by  $f$  from any choice set is equal to the choice implied by  $c$ . First, note that this holds for  $X$ , since by [Observation 2](#),  $f(a, X, X, U^A) = f(b, X, X, U^A)$

for every  $a, b \in X$  and every  $A$  with which there is an IIA violation associated. Moreover,  $f(c(X), X, X, u^*) > f(a, X, X, u^*) \forall a \in X \setminus \{c(X)\}$  by P2. Then repeated application of P3 implies  $f(c(X), X, X, U_c) > f(a, X, X, U_c) \forall a \in X \setminus \{c(X)\}$ .

Next, consider any  $A \subsetneq X$  which causes an IIA violation. Suppose  $A \in I_j$ . **Observation 2** implies that for any  $B \in (\bigcup_{l=1}^j I_l) \setminus A$ ,  $f(a, A, U^B) = f(a', A, U^B) \forall a, a' \in A$ , and  $f(c(A), A, X, U^A) > f(a, A, X, U^A) \forall a \in A$ . Then repeated implication of P3 implies  $f(c(A), A, X, U_j) > f(a, A, X, U_j) \forall a \in A$ . By construction then  $f(c(A), A, X, U_c) > f(a, A, X, U_c) \forall a \in A$ .

There are three cases to check for a set  $A$  that does not cause an IIA violation.

*Case 1:* For all  $a \in A$ , there is no  $B \supset A$  such that  $a = c(B)$ . Then by construction  $u^*(c(B)) > u^*(b) \forall b \in B \setminus \{c(B)\}$ . Moreover, by **Observation 2**,  $f(b, B, X, U^A) = f(b', B, X, U^A) \forall b, b' \in B$  and  $A$  with which an IIA violation is associated. Repeated use of P3, together with P2, implies  $f(c(B), B, X, U_c) > f(b, B, X, U_c) \forall b \in B$ .

*Case 2:* There is a unique  $a \in A$  such that for some  $B \supset A$ ,  $c(B) = a$ . First we note that  $a = c(A)$  is necessary, otherwise  $A$  would have caused an IIA violation. There are two subcases:

*Case 2a:* For every  $B$  such that  $B \supset A$  and  $c(B) = a$ ,  $B$  did not cause an IIA violation. This means that for all  $B \supset A$ ,  $c(B) \notin A \setminus \{c(A)\}$ . So just like in Case 1,  $u^*(c(B)) > u^*(b) \forall b \in B \setminus \{c(B)\}$ , and  $f(b, B, X, U^A) = f(b', B, X, U^A) \forall b, b' \in B$  and  $A$  with which an IIA violation is associated. Hence,  $f(c(B), B, X, U_c) > f(b, B, X, U_c) \forall b \in B$ .

*Case 2b:* There is  $B \supset A$  with  $c(B) = a$  such that  $B$  caused an IIA violation. Consider any smallest such  $B$ , and suppose  $B \in I_j$ . By **Observation 2**, for any  $A \in \bigcup_{l=1}^j I_l$  either  $f(c(B), B, X, U^A) > f(b, B, X, U^A) \forall b \in B$ , or  $f(b, B, X, U^A) = f(b', B, X, U^A) \forall b, b' \in B$ . But then repeated application of P3 implies that  $f(c(B), B, X, U_j) > f(b, B, X, U_j) \forall b \in B$ . By construction,  $f(c(B), B, X, U_c) > f(b, B, X, U_c) \forall b \in B$ .

*Case 3:* There exist at least two elements in  $A$  that have each been chosen in some superset. First, note that one of those elements must be  $a = c(A)$ , otherwise  $A$  would have caused an IIA violation. Let  $\{b_i\}_i$  be the set of elements other than  $a$  such that  $b_i \in A$  and  $b_i = c(B_i)$  for some  $B_i \supset A$ . Drop any  $b_i$ 's such that  $B_i \subset B_m$  for some  $m$  and call the remaining set  $\{b_j\}$ . Because  $A$  did not cause an IIA violation by assumption, it must be that for each  $b_j$  there is  $A'_j$  such that  $A \subset A'_j \subset B_j$  and  $c(A'_j) \in A$ . Because  $B_j$  does not contain any  $B_k$ , we know  $c(A'_j) = a$ . For each  $j$  there may be multiple such  $A'_j$ 's; consider only the maximal  $A'_j$  with respect to the minimal  $B_j$ . Now by maximality, for any  $A''$  such that  $A'_j \subset A'' \subset B_j$ ,  $c(A'') \notin A$ . If there is  $A''$  such that  $c(A'') \in A'_j$ , since  $c(A'') \neq a$ , by definition  $A'_j$  caused an IIA violation with respect to the first such  $A''$ . If for every  $A''$  it is the case that  $c(A'') \notin A'_j$ , then once again  $A'_j$  caused an IIA violation with respect to  $B$ . Either way, since  $c(A'_j) = a$ , we added selves to ensure this choice for every  $j$ . This means

that  $a$  should be the choice from  $A$  unless for some set  $B'$  between the smallest-sized  $A'_j$  and  $A$  we have  $c(B') \in A \setminus \{a\}$  and selves were added. But by minimality of the  $B_j$ 's there cannot be such a set. ■

### Proofs of Theorem 2 and Theorem 3

Let  $X = \{a, b, c\}$ . For compactness, we use the notation

$$\begin{aligned} x_1 &= f(a, \{a, b, c\}, X, U) - f(b, \{a, b, c\}, X, U), \\ x_2 &= f(b, \{a, b, c\}, X, U) - f(c, \{a, b, c\}, X, U), \\ x_3 &= f(a, \{a, c\}, X, U) - f(c, \{a, c\}, X, U), \\ x_4 &= f(b, \{b, c\}, X, U) - f(c, \{b, c\}, X, U), \\ x_5 &= f(a, \{a, b\}, X, U) - f(b, \{a, b\}, X, U). \end{aligned}$$

We first prove two lemmas.

**Lemma 1.** *If  $x_3 \neq x_4 + x_5$ , and if any one of the three equations  $2x_1 + x_2 - x_3 - x_5 = 0$ ,  $x_1 + 2x_2 - x_3 - x_4 = 0$ , or  $x_1 - x_2 + x_4 - x_5 = 0$  fails, then the aggregator is triple-solvable (with  $k_f$  at most  $2 + 3|U|$ ).*

*Proof.* The first column in the table lists the aggregate values for selves  $U$ . But by neutrality, we know that if we can generate the values in column 1, we can also generate the values in the 2nd column using the permutation  $(bc)(a)$  over the alternatives, generate the values in the 3rd column using the permutation  $(ab)(c)$  over the alternatives, and so on. By using duplication to evaluate each of the values  $f \circ u$  and  $f \circ u'$  each generated by a single self  $u$  and  $u'$ , with the rankings given in the 6th and 7th headers, respectively, we can also generate the values in those respective columns.

1 : $U$	2 : $(bc)(a)$	3 : $(ab)(c)$	4 : $(abc)$	5 : $(acb)$	6 : $a \sim b \succ c$	7 : $a \succ b \sim c$
$x_1$	$x_1 + x_2$	$-x_1$	$x_2$	$-x_1 - x_2$	0	$x_1$
$x_2$	$-x_2$	$x_1 + x_2$	$-x_1 - x_2$	$x_1$	$x_1$	0
$x_3$	$x_5$	$x_4$	$-x_5$	$-x_4$	$x_1$	$x_1$
$x_4$	$-x_4$	$x_3$	$-x_3$	$x_5$	$x_1$	0
$x_5$	$x_3$	$-x_5$	$x_4$	$-x_3$	0	$x_1$

Then, determinants of three possible  $5 \times 5$  matrices, each composed of five of the columns above,

may be calculated to obtain:

$$\begin{aligned}\text{Det}(1|3|5|6|7) &= x_1^2(x_1 + 2x_2 - x_3 - x_4)(2x_1 + x_2 - x_3 - x_5)(x_3 - x_4 - x_5), \\ \text{Det}(1|2|5|6|7) &= x_1^2(2x_1 + x_2 - x_3 - x_5)(x_3 - x_4 - x_5)(x_1 - x_2 + x_4 - x_5), \\ \text{Det}(2|3|4|6|7) &= -x_1^2(x_1 + 2x_2 - x_3 - x_4)(x_3 - x_4 - x_5)(x_1 - x_2 + x_4 - x_5).\end{aligned}$$

To prove the result, it suffices to show that there exists  $U$  such that defining  $x_1, x_2, \dots, x_5$  as above, one of the determinants above must be nonzero. If one of those determinants is nonzero, then we have find a vector  $(c_1, c_2, c_3, c_4, c_5)$  such that the nonsingular matrix times  $(c_1, c_2, c_3, c_4, c_5)$  is equal to  $(0, 0, 0, 0, \beta)$  for some  $\beta \neq 0$ . Using scaling, each  $c_i$  can be pulled in so that the  $U$  corresponding to the  $i$ -th column is multiplied by  $c_i$ . The resulting set of selves provides a triple-basis (and therefore we can get triple solvability through scaling that triple-basis).

The proof is completed in light of the linear dependence of the equations  $2x_1 + x_2 - x_3 - x_5 = 0$ ,  $x_1 + 2x_2 - x_3 - x_4 = 0$ , and  $x_1 - x_2 + x_4 - x_5 = 0$ : if any one of these fails, there must be a second which fails too. ■

**Lemma 2.** *Suppose there exists  $U \in \mathcal{U}(\{a, b, c\})$  such that  $x_3 \neq x_4 + x_5$  and  $f \circ U$  rationalizes choice where the worst element in the transitive pairwise ranking is best in the triple. Then either  $2x_1 + x_2 \neq x_3 + x_5$  or  $x_1 + 2x_2 \neq x_3 + x_4$ .<sup>25</sup>*

*Proof.* By neutrality and symmetry of the condition  $x_3 - x_4 - x_5 \neq 0$ , there are two types of choice behaviors we must examine to prove the result:

*Case 1:*  $a \succ_P b \succ_P c$  on the pairs, and  $c \succ_T b \succeq_T a$  on the triple. That is,  $x_3, x_4, x_5 > 0$ , with  $x_1 \leq 0$  and  $x_2 < 0$ . But then  $2x_1 + x_2 \neq x_3 + x_5$ , since the LHS is negative and the RHS is positive.

*Case 2:*  $a \succ_P b \succ_P c$  on the pairs, and  $c \succ_T a \succeq_T b$  on the triple. That is,  $x_3, x_4, x_5 > 0$ , with  $x_1 \geq 0$ ,  $x_2 < 0$ . If we can find  $U$  such that  $f \circ U$  rationalizes this behavior, then observe that  $x_1 + 2x_2$  is negative. Hence  $x_1 + 2x_2 \neq x_3 + x_4$  because the RHS is positive. ■

**Theorem 3** follows as a corollary of **Lemma 1**. To complete the proof of **Theorem 2**, fix an aggregator  $f$  and suppose there is  $U \in \mathcal{U}(\{a, b, c\})$  such that  $f \circ U$  can rationalize third-place choice, and that there is  $U' \in \mathcal{U}(\{a, b, c\})$  such that  $f \circ U'$  can rationalize intransitive behavior. If  $f \circ U$  satisfies  $x_3 \neq x_4 + x_5$  (for this collection of selves) then the proof is complete by **Lemma 1** and **Lemma 2**. Otherwise, consider the collection of selves  $(U, \varepsilon \cdot U')$ , where  $\varepsilon > 0$  is a positive scalar, and small enough that by P4 we still rationalize third-place choice with  $f \circ (U, \varepsilon \cdot U')$ . At

---

<sup>25</sup>The above is also true for one type of second-best choice from the triple:  $a \succ_P b \succ_P c$  on the pairs, and  $b \succ_T c \succeq_T a$  on the triple. If there is  $U$  such that  $f \circ U$  rationalizes this behavior, then  $x_3, x_4, x_5 > 0$  and  $x_1 \leq 0$ ,  $x_2 > 0$ . Observe that  $2x_1 + x_2 < 0$  since this is  $f_a(U) - f_b(U) + f_a(U) - f_c(U)$ . Therefore,  $2x_1 + x_2 \neq x_3 + x_5$ , as the RHS is positive.

the same time, note that  $f \circ U'$  must satisfy  $x_3 \neq x_4 + x_5$  (for this collection of selves) since it rationalizes intransitive behavior. Hence,  $f \circ (U, \varepsilon \cdot U')$  also satisfies  $x_3 \neq x_4 + x_5$ . Hence [Lemma 1](#) and [Lemma 2](#) apply. ■

## References

- Arrow, Kenneth J. and Hervé Raynaud**, *Social Choice and Multicriterion Decision-Making*, Cambridge, Massachusetts: The MIT Press, 1986.
- Benabou, Roland and Marek Pycia**, “Dynamic Inconsistency and Self-Control: A PlannerDoer Interpretation,” *Economics Letters*, 2002, 77, 419-424.
- Bernheim, Douglas and Antonio Rangel**, “Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics,” *Working Paper*, 2007.
- Browning, Martin and Pierre-André Chiappori**, “Efficient Intra-Household Allocations: A General Characterization and Empirical Tests,” *Econometrica*, 1998, 66, 1241–1278.
- Chambers, Chris and Takashi Hayashi**, “Choice and Individual Welfare,” *Working Paper*, 2008.
- Chatterjee, Kalyan and R. Vijay Krishna**, “Menu Choice, Environmental Cues and Temptation: A ‘Dual Self’ Approach to Self-Control,” *American Economic Journal: Microeconomics*, forthcoming.
- Cherepanov, Vadim, Tim Feddersen, and Alvaro Sandroni**, “Rationalization,” *Working paper*, 2008.
- Chiappori, Pierre-André and Ivar Ekeland**, “The Microeconomics of Group Behavior: General Characterization,” *Journal of Economic Theory*, 2006, 130, 1–26.
- Debreu, Gérard**, “Topological Methods in Cardinal Utility Theory,” *Cowles Foundation Paper 156*, 1959.
- Dekel, Eddie, Barton Lipman, and Aldo Rustichini**, “A Unique Subjective State Space for Unforeseen Contingencies,” *Econometrica*, 2001, 69, 891–934.
- Dhillon, Amrita and Jean-Francois Mertens**, “Relative Utilitarianism,” *Econometrica*, 1999, 67, 471–498.
- Donahue, Eileen, Richard Robins, Brent Roberts, and Oliver John**, “The Divided Self: Concurrent and Longitudinal Effects of Psychological Adjustment and Social Roles on Self-Concept Differentiation,” *Journal of Personality and Social Psychology*, 1993, 64, 834–846.
- Evren, Ozgur and Efe Ok**, “On the Multi-Utility Representation of Preference Relations,” *Working Paper*, 2007.
- Fudenberg, Drew and David Levine**, “A Dual Self Model of Impulse Control,” *American Economic Review*, 2006, 96, 1449–1476.
- Green, Jerry and Daniel Hojman**, “Choice, Rationality, and Welfare Measurement,” *Working Paper*, 2009.

- Gul, Faruk and Wolfgang Pesendorfer**, “Temptation and Self-Control,” *Econometrica*, 2001, 69, 1403-1436.
- Kahneman, Daniel, Peter Wakker, and Rakesh Sarin**, “Back to Bentham? Explorations of Experienced Utility,” *The Quarterly Journal of Economics*, 1997, 112, 375–405.
- Kalai, Gil, Ariel Rubinstein, and Ran Spiegler**, “Rationalizing Choice Functions By Multiple Rationales,” *Econometrica*, 2002, 70, 2481-2488.
- Kamenica, Emir**, “Contextual Inference in Markets: On the Informational Content of Product Lines,” *American Economic Review*, 2008, 98, 2127–2149.
- Karni, Edi**, “Impartiality: Definition and Representation,” *Econometrica*, 1998, 66, 1405–1415.
- Keeney, Ralph L. and Howard Raiffa**, *Decisions with Multiple Objectives*, Cambridge, U.K.: Cambridge University Press, 1993.
- Kivetz, Ran, Oded Netzer, and V. Srinivasan**, “Alternative Models for Capturing the Compromise Effect,” *Journal of Marketing Research*, 2004, 41, 237–257.
- Kochov, Asen**, “The Epistemic Value of a Menu and Subjective States,” *Working Paper*, 2007.
- Kreps, David**, “A Representation Theorem for ‘Preference for Flexibility’,” *Econometrica*, 1979.
- Lachmann, Frank M.**, “How Many Selves Make a Person?,” *Contemporary Psychoanalysis*, 1996, 32, 595–614.
- Lee, Leonard, On Amir, and Dan Ariely**, “In Search of Homo Economicus: Preference Consistency, Emotion, and Cognition,” *Working Paper*, 2007.
- Manzini, Paola and Marco Mariotti**, “Sequentially Rationalizable Choice,” *American Economic Review*, 2007, 97, 1824–1839.
- Masatlioglu, Yusufcan and Daisuke Nakajima**, “Theory of Choice By Elimination,” *Working Paper*, 2007.
- and **Efe Ok**, “Rational Choice with Status Quo Bias,” *Journal of Economic Theory*, 2005, pp. 1–29.
- Maskin, Eric**, “A Theorem on Utilitarianism,” *The Review of Economic Studies*, 1978, pp. 93–96.
- May, Kenneth O.**, “Intransitivity, Utility, and the Aggregation of Preference Patterns,” *Econometrica*, 1954, pp. 1–13.
- Mitchell, Stephen**, *Hope and Dread in Psychoanalysis*, New York: BasicBooks, 1993.
- Power, M.J.**, “The Multistory Self: Why the Self is More Than the Sum of Its Autoparts,” *Journal of Clinical Psychology*, 2007, 63, 187–198.
- Saari, Donald G.**, “Explaining All Three-Alternative Voting Outcomes,” *Journal of Economic Theory*, 1999, 87, 313–355.

- Salant, Yuval**, “Procedural Analysis of Choice Rules with Applications to Bounded Rationality,” *Working Paper*, 2007.
- and **Ariel Rubinstein**, “Choice with Frames,” *The Review of Economic Studies*, 2008, *75*, 1287.
- Segal, Uzi**, “Let’s Agree That All Dictatorships Are Equally Bad,” *Journal of Political Economy*, 2000, *108*, 569–589.
- Sen, Amartya K.**, “Choice Functions and Revealed Preference,” *The Review of Economic Studies*, 1971, *38*, 307–317.
- , “Internal Consistency of Choice,” *Econometrica*, 1993, *61*, 495–521.
- Shafir, Eldar, Itamar Simonson, and Amos Tversky**, “Reason-Based Choice,” *Cognition*, 1993, *49*, 11–36.
- Simonson, Itamar**, “Choice Based on Reasons: The Case of Attraction and Compromise Effects,” *Journal of Consumer Research*, 1989, *16*, 158–174.
- and **Amos Tversky**, “Choice in Context: Tradeoff Contrast and Extremeness Aversion,” *Journal of Marketing Research*, 1992, pp. 281–295.
- Strotz, R.H.**, “Myopia and Inconsistency in Dynamic Utility Maximization,” *Review of Economic Studies*, 1955, *23*, 165–180.
- Tversky, Amos**, “Intransitivity of Preferences,” *Psychological Review*, 1969, *76*, 31–48.
- and **Daniel Kahneman**, “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *Quarterly Journal of Economics*, 1991, *106*, 1039–1061.
- and **Itamar Simonson**, “Context-Dependent Preferences,” *Management Science*, 1993, *39*, 1179–1189.