Rationalizing Choice with Multi-Self Models*†

Attila Ambrus‡ Duke University

Kareen Rozen§ Yale University

This version: July 2013

Abstract

This paper shows that in situations in which the preferences of multiple selves are aggregated into a collective decision, even if the researcher has a fully specified theory of how preferences get aggregated, there are typically no testable implications of the theory unless there is an a priori restriction on the number of selves. This result has implications in both interpersonal and intrapersonal decision-making, calling attention to the importance of collecting reliable information on the number of selves (motivations, in the interpersonal context) participating in the decision. We establish our main result by finding a linear relationship between the number of selves and the set of choice functions that a given aggregator is guaranteed to rationalize with the given number of selves. The latter set is connected to the number of IIA violations implied by the choice function, a new measure of the amount of irrationality a choice function exhibits.

JEL Codes: D11, D13, D71
Keywords: Collective decision-making, Multiple selves, Testable implications, IIA violations, Rationalizability

*First version: April 2008. Early versions were distributed under the title "Revealed Conflicting Preferences."

†We are grateful to Geoffroy de Clippel, 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, Philipp Sadowski, Larry Samuelson, Rani Spiegler, Tomasz Strzalecki, and especially the editor and referees for valuable comments and suggestions. We also thank seminar audiences at Brown, Harvard, MIT, Montreal, NYU, UCL, Yale and the North American Summer Meeting of the Econometric Society.

‡Address: Dept. of Economics, Duke University, 419 Chapel Drive, Box 90097, Durham, NC 27708. E-mail: aa231@duke.edu.

§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. Home page: http://www.econ.yale.edu/~kr287/.

1 Introduction

Since the seminal work of May (1954), a growing number of papers have proposed models of multi-self decision-making, with the primary motivation of accommodating context-dependent behavior (thereby relaxing the axiom of *Independence of Irrelevant Alternatives*, or IIA, which requires that if an alternative is chosen from a set, it is also chosen from any subset in which it is contained).[1,2] Formal models of multi-self decision-making include among others Kalai, Rubinstein and Spiegler (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. A parallel literature, including for example Apps and Rees (1988), Chiappori (1988), Browning and Chiappori (1998) and Cherchye, De Rock and Vermeulen (2007), studies interpersonal aggregation of preferences in a household or a larger community.

Many papers in the literature assume a particular method of aggregating preferences, but do not put a priori restrictions on the number of selves involved in the decision. Each of these practices is potentially justified. In interpersonal contexts, the decision-making procedure can be observable, or suggested by theoretical considerations. In intrapersonal contexts, theoretical considerations, experimental data, or neuroscience research can suggest a certain method of preference aggregation for different selves, or motivations, of the individual. At the same time, the researcher might not know the number of selves relevant for the decision; such data limitations include the possibility of unobserved selves, or at the extreme, having no available data on the number of selves.

In this paper, we examine whether specifying a certain method of preference aggregation generates testable predictions on choice behavior without putting an a priori restriction on the number of selves. For this, we need to know what choice behaviors can be explained by a given number of selves. For some aggregators, this is easy to determine. For example, if the decision maker (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)?

In order to investigate this question, we lay out in Section 2 a framework that incorporates various models of multi-self decision-making which have been proposed in the literature. In particular,

[1] The IIA condition is also known as Sen's α (Sen 1971) and for single-valued choice, is equivalent to being able to describe choices as the maximization of a strict, complete and transitive preference.

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

we model the DM or group as a collection of selves (of possibly different types) and an *aggregation rule* f (decision-making method) which combines the selves' utility functions in a possibly context-dependent way. That is, given a choice set A, and selves S, an aggregator f specifies an aggregate utility for every alternative in A. Each aggregator in the framework captures a particular theory of multi-self decision-making. We examine a broad class of aggregators characterized by five simple properties 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 satisfying our axioms. Since our results apply for a broad class of multi-self models, we provide a meta-analysis of various models proposed in the literature, and offer a way to characterize the explanatory power of such models.

An important feature of the set of aggregators that we focus on is that aggregation can depend on cardinal information in the selves' utilities. This is partly motivated by the fact that many existing models of multi-self decision-making make use of cardinal information embedded in different selves' utility functions. Furthermore, the use of cardinal information (intensity of preferences) is natural to assume in intrapersonal decision-making, and in certain interpersonal decision-making situations as well, such as household decisions. A second feature of the aggregators we consider, related to cardinality, is the possibility of compromise among selves. As opposed to the models provided in Kalai et al. (2002) and Cherepanov, Feddersen and Sandroni (forthcoming), 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 the primary rationales for purchasing may depend on the set of available products.

Our primary goal is to investigate the behaviors a model of multi-self decision-making can *rationalize* (explain). 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 S if she selects the unique maximizer of the selves' aggregate utility from every choice set. The DM's choice behavior need not satisfy IIA. In Section 3 we define a measure of a choice function's irrationality: the number of IIA violations it exhibits.

Our main results, in Section 4 and Section 5, 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. For example, in the important class of scale-invariant aggregators (for which the unit of utility measurement does not change the ordinal rankings in the aggregation), 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. Furthermore, we show that in a formal sense, aggregators satisfying the above property are generic; therefore, one generically cannot have testable predictions without restricting the number of selves. 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 rationalizable behaviors. This need not be the case outside the class of models studied here: Manzini and Mariotti (2007) and de Clippel and Eliaz (2012), for example, have shown that their models can explain only certain types of irrational behaviors, even when using arbitrarily many rationales. We establish our theorems by finding a simple linear relationship between the number of selves in the model, and the number of IIA violations the choice function can have while still guaranteeing that it can be rationalized. Our results can be seen as drawing 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.[3]

Our results differ from Kalai et al. (2002), who examine the complexity of a rationalization as a function of the number of alternatives available. They say a collection of preference orderings rationalizes a choice function if the choice from each set is optimal for some preference, and show it suffices to posit as many selves as there are alternatives to explain any behavior. To rationalize a choice function, they assign each utility function the sets over which it acts as dictator, which amounts to modifying the method of aggregation. By contrast, this paper studies the set of behaviors rationalizable by a fixed aggregator. Our results also differ from those in the household choice literature, such as Chiappori (1988), Browning and Chiappori (1998), Chiappori and Ekeland (2006), and Cherchye, De Rock and Vermeulen (2007, 2009, 2011). These works focus on rationalizing demand in a market environment. The results they obtain are nonparametric in the sense that they do not rely on the particular functional specification chosen for the preferences or for the intra-household allocation process. Our finite (though abstract) choice setting is closer in spirit to that of Cherchye, De Rock and Vermeulen, who study a global revealed preference framework assuming a finite set of demand and price observations. Green and Hojman (2009) 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 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.[4] The rest of their paper focuses on welfare analysis.

[3]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).

[4]This paper's result on rationalization is independent of their monotonicity theorem.

2 Framework

We observe a collective choice behavior on a finite set of alternatives X. Denote by $P(X)$ the set of nonempty subsets of X. The *collective choice function* $c : P(X) \to X$ identifies the alternative $c(A) \in A$ chosen from each $A \in P(X)$. A *rationalization* of the collective choice function consists of a collection of *selves* and a *model of aggregation* that combines the utilities of different selves in a possibly menu-dependent way into an aggregate utility function. In an interpersonal context, selves represent different individuals. In an intrapersonal context, selves represent the decision maker's conflicting motivations or priorities. The aggregator corresponds to a method of "sorting out" priorities of different selves to come to a decision.

In order for our framework to encompass as many of the multi-self models proposed in the existing literature as possible, we permit selves to have "types" and consider potentially asymmetric aggregators that treat selves differently according to their type. Formally, a model of aggregation (f, T) specifies a set T of the possible types a self may take, and a function f that gives the aggregate utility for every alternative a in every choice set A, for any (finite) grand set of alternatives X and any collection of selves defined over X and T. A single self s is given by a pair (u, t), where $u : X \to \mathbb{R}$ is a utility function and $t \in T$ is the self's type. Hence, each self is an element of $\mathbb{R}^X \times T$. A collection of selves S is an unordered list of selves.[5] Formally, for a given grand set of alternatives X and set of possible types T, a collection of selves S is an element of $\mathcal{S}(X, T) = \cup_{n=1}^{\infty} \mathcal{S}^n(X, T)$, where $\mathcal{S}^n(X, T)$ is the set of all unordered lists of selves over X that contain n elements. We denote the number of selves in a particular collection S by $|S|$, or simply n when no confusion would arise.[6]

The aggregator f specifies an aggregate utility for every alternative a in every choice set A, given any (finite) grand set of alternatives X, set of types T, and collection of selves S. Formally, the domain over which f is defined is $\{a, A, S, X, T | X \in \mathcal{X}, S \in \mathcal{S}(X, T), A \in P(X), a \in A\}$, 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.[7] An aggregation rule may be seen as a particular theory of how selves are activated by choice sets: the aggregator determines the weight each self receives on the choice set as a function of its utility levels over the alternatives. Formally, the grand set of alternatives X is an argument of the aggregator, not only because the evaluation of an alternative $a \in A$ might depend on alternatives outside the choice 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. For simplicity, we

[5] In combinatorics this object is also referred to as a multiset.

[6] 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 S, then so does any increasing transformation of f. Similarly, given any representation S and f, one can obtain an equivalent representation by applying a monotone transformation of utilities in S, if a corresponding transformation is applied to the aggregation function f as well.

[7] We can permit aggregators with restricted domains: let $\hat{\mathbb{R}}^X$ be a convex subset of $\mathbb{R}^X$ and let $\mathcal{S}^n = (\times_{i=1}^{n} \hat{\mathbb{R}}^X) \times T$.

will suppress notational dependence of f on X and T, writing simply $f(a, A, S)$, whenever doing so would not cause confusion.

Given a model, we say that a collection of selves rationalizes a choice function if from every choice set, the alternative that maximizes the aggregated utility is precisely the one selected by the choice function.[8] Note that this definition requires a unique maximizer of aggregate utility.

Definition 1. *A model (f, T) rationalizes a choice function $c(\cdot)$ on X if there exists a finite collection of selves $S \in \mathcal{S}(X, T)$ such that for every $A \in P(X)$, $c(A) = \arg\max_{a \in A} f(a, A, S)$.*

2.1 The class of models studied

We study a class $\mathcal{F}$ of models of multi-self aggregation satisfying the following properties, most of which are familiar from the theory of social choice. These properties are satisfied by several previously proposed multi-self models. In the resulting class of models, aggregation of utilities is cardinal and the framing effect of a choice set operates only through the utility levels of the different selves. Before introducing these properties, it will be useful to define the following notation. For any collections of selves $S = \langle s_1, \ldots, s_{|S|} \rangle$ and $S' = \langle s'_1, \ldots, s'_{|S'|} \rangle$ in $\mathcal{S}(X, T)$, we denote by $\langle S, S' \rangle$ the combined collection $(s_1, \ldots, s_{|S|}, s'_1, \ldots, s'_{|S'|}) \in \mathcal{S}(X, T)$.

P1 (Neutrality)**.** *For any permutation $\pi : X \to X$, $f(a, A, S) = f(\pi(a), \pi(A), \langle (u \circ \pi^{-1}, t) \rangle_{(u,t) \in S})$.*

P2 (Consistency)**.** *For any $s = (u, t)$, $u(a) \geq u(b)$ if and only if $f(a, A, s) \geq f(b, A, s)$.*

P3 (Reinforcement)**.** *If both $f(a, A, S) \geq f(b, A, S)$ and $f(a, A, S') \geq f(b, A, S')$ then $f(a, A, \langle S, S' \rangle) \geq f(b, A, X, \langle S, S' \rangle)$, with strict inequality if one of the above is strict.*

P4 (Continuity to near-indifferent additions)**.** *If $f(a, A, S) > f(b, A, S)$, then for any $k \in Z$ there exists $\delta > 0$ such that $f(a, A, \langle S, S' \rangle) > f(b, A, \langle S, S' \rangle)$ for any $S' \in \mathcal{S}^k(X)$ with the property that $\max_{a,b \in A, A \subseteq X, s' \in S'} |f(a, A, s') - f(b, A, s')| < \delta$.*

P5 (Profile equivalence)**.** *If $u(a) = u(a')$ for all $(u, t) \in S$ then $f(b, A \cup \{a\}, S) = f(b, A \cup \{a'\}, S)$ for all $b \in A$.*

While these properties are not without loss of generality, they are satisfied by many multi-self models that have been proposed in the literature. Neutrality implies that the names of alternatives do not affect their ranking (only utilities affect rankings). Consistency requires respecting the preference of a lone self. Reinforcement requires that if two separate collections of selves S and S' each

[8] We note that an aggregator f encodes additional information, such as the ranking of unchosen alternatives in each set, that might be observable using a larger data set than that provided by a choice function. However, using only simple revealed preference on the choice from a menu, only the best choice from each set (i.e., the choice function) is elicited in light of the potential menu-dependence of choices.

prefer the alternative a to the alternative b, then the combined collection of selves, obtained by merging collections of selves S' and S, also prefers a to b. Consistency and reinforcement together imply Pareto-optimality. Continuity to near-indifferent additions introduces a cardinal feature into the method of aggregation. It does not require that f (or the ordering of the alternatives implied by f) be continuous in the utilities of selves, for a fixed number of selves; it only requires that if a collection of selves leads to a strict aggregate preference for a over b, then that preference is not overturned when adding selves for which the aggregate utility difference between alternatives is sufficiently small. That is, preference intensity matters. In view of consistency and reinforcement, assuming P4 is weaker than assuming full continuity.[9] Finally, profile equivalence says that aggregation is only affected by the set of available 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. This means that adding "duplicate" elements to a set, which replicate the exact utility levels of some element already in the set, does not affect the rankings of alternatives. However, increasing the size of a set can still affect the decision-maker when the new elements change the set of possible utility levels.

For ease of exposition, in the main text we also restrict attention to aggregators for which the aggregate utility of an alternative in a choice set A is independent of alternatives outside of A. (See Supplementary Appendix D for an extension of our results without imposing this assumption).

P6 (Independence of unavailable alternatives)**.** *Let X, X' be two grand sets of alternatives and consider any $A \subseteq X \cap X'$. Take any collection of types $(t_1, \ldots, t_n)$ and any two collections $(u_1, \ldots, u_n)$ and $(u'_1, \ldots, u'_n)$ of utility functions over X and X', respectively. If $u_i(x) = u'_i(x)$ for each $x \in A$ and each i, then the aggregator satisfies $f(\cdot, A, \langle (u_i, t_i) \rangle_i, X, T) = f(\cdot, A, \langle (u'_i, t_i) \rangle_i, X', T)$.*

2.2 Examples of aggregators

The following are examples of context-dependent aggregators satisfying P1-P6, that are equivalent or closely related to models proposed in the existing literature. In the first four examples, the aggregator treats all types symmetrically, so we may take the type set T to be a singleton.

Example 1 - Utilitarianism. The aggregate utility of an alternative a in a choice set A is given by $\sum_{(u,t)\in S} u(a)$. Note that the utility of an alternative is independent of the choice set within which it is evaluated.

Example 2 - Generalization of Tversky (1969). The aggregate utility of an alternative a in a choice set A is $\sum_{(u,t)\in S} \Phi(\max_{b\in A} u(b) - \min_{b\in A} u(b))u(a)$, where the contribution of a self to the aggregate utility depends via Φ on the range of u over choice set A. For binary choice sets, this

[9] P2 means that a fully indifferent self leads to aggregate indifference, and iterating this and using P3 means that adding a finite number of indifferent selves doesn't affect an existing strict preference; if we were to assume full continuity on top of this, P4 would be implied.

reduces to the *additive difference model* of Tversky (1969), which was proposed to explain intransitive pairwise choice through the aggregation of criterion-by-criterion comparisons of alternatives.[10] If Φ in that model is increasing, utility functions with a greater intensity of preference over the set A receive greater weight in the aggregate utility. The case $\Phi(x) = x$ is Kőszegi and Szeidl (2012)'s *focus-weighted* model. If Φ is decreasing, the model may be seen as 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.

Example 3 - Nash bargaining solution with an endogenous disagreement point. The aggregate utility of an alternative a in a choice set A is $\prod_{(u,t)\in S}(\kappa + u(a) - \min_{a'\in A} u(a'))$, where κ is any positive constant to ensure each term is strictly positive.

This example, which specifies the worst outcome as the disagreement point, is similar to Kaneko and Nakamura (1979), although they assume the utility of the worst outcome is the same in all choice sets. A more general theory of context-dependent disagreement points in the bargaining solution is offered by Conley, McLean and Wilkie (1997).

Example 4 - Loss aversion of Tversky and Kahneman (1991), with endogenous reference point. The aggregate utility of an alternative a in a choice set A is given by $\sum_{(u,t)\in S} m\big(u(a)\big) + \sum_{u\in U} \ell\Big(u(a) - r(\{u(a')\}_{a'\in A})\Big)$, where $r(\cdot)$ determines the reference point against which $u(a)$ is evaluated; $m(\cdot)$ captures the impact of absolute valuations on aggregate utility; and the loss aversion function $\ell(\cdot)$ satisfies the properties proposed by Tversky and Kahneman (1991): steeper disutility from losses than utility from gains, and weakly diminishing sensitivity.

The above model has been applied in various forms. In Orhun (2009), each u can be interpreted as the valuation of alternatives under some attribute. Orhun (2009) finds the optimal product line for a model corresponding to the case where m is linear, ℓ is the standard kinked-linear loss aversion function (that is, $\ell(x) = x$ for $x > 0$, $\ell(x) = \lambda x$ for $x < 0$ and some $\lambda > 1$), and r is a weighted average of valuations. Kivetz et al. (2004) consider goods (e.g., laptops) which have defined attribute levels (e.g., processor speed) and posit utility levels ("partworths") for a given attribute. Their *contextual concavity model* specifies $r(\cdot) \equiv \min(\cdot)$, $m(\cdot) \equiv 0$, and $\ell(\cdot) \equiv (\cdot)^\rho$ for some concavity parameter ρ. They also introduce a *type-dependent* version of their model, where the concavity parameter ρ depends on the type of attribute to which the self corresponds.

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^{lr} (the expected present value of the utility stream induced by the choice in the present), and

[10] Tversky (1969) accounts for intransitive pairwise choice behavior by positing utilities $v_1, v_2, \ldots, v_n$ and an odd $\phi : \mathbb{R} \to \mathbb{R}$ such that $x \succ y$ if and only if $\sum_{i=1}^n \phi(v_i(x_i) - v_i(y_i)) > 0$. Observe that a is preferred to b in the pair $\{a, b\}$ if and only if $\sum_{u\in U} \Phi(|u(a) - u(b)|)(u(a) - u(b))$, where each summand is an odd function of $u(a) - u(b)$.

the short-run self, with utility function u^{sr} (the present period consumption utility).[11] Using our terminology, there are two types of selves, long run (lr) and short run (sr), and their reduced form representation assigns to alternative a the aggregate utility $u^{lr}(a) - C(a)$, 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, $C(a) = \gamma\Big(\max_{a' \in A} u^{sr}(a') - u^{sr}(a)\Big)^{\psi}$. More generally, there may be multiple long-run considerations and multiple short-run temptations.

Example 5 - Costly self-control aggregators. The set of possible types is $T = \{lr, sr\}$ and the aggregate utility of an alternative a in a choice set A is $f(a, A, S) = \sum_{(u,lr)\in S} u(a) - \sum_{(u,sr)\in S} \gamma\Big(\max_{a' \in A} u(a') - u(a)\Big)^{\psi}$. Of course, given the utilitarian aggregation of the long-run selves, they could equivalently be represented using a single utility function.

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. 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

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

In light of $c(\{a, b, c, d, e, f\}) = d$, IIA dictates that the last three choices should be d (but they are

[11] 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^{lr} is not restricted by the short-run utility u^{sr}. If continuation values cannot be arbitrary (for example they have to be nonnegative) then u^{sr} restricts the possible values of u^{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.

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 $c(\{a,b,c,d\}) = b$. The idea is that one "resets" the point from which the IIA implication must hold: if $c(B)$ is chosen from B but is not chosen from $A \subset B$, then for all subsets of A in which $c(A)$ is contained, one expects $c(A)$ to continue being chosen. With this idea in mind, our accounting procedure counts the number of such resets, associating an IIA violation with a choice set when it is the largest set whose choice violates the IIA implication coming from a superset.

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 sketch of proof for our main result in Section 5 illustrates the connection between this definition and our rationalization procedure. 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.[12]

4 Scale-invariant models

We begin by introducing our results for a special class $\mathcal{F}^*$ of type-independent aggregators satisfying P1-P6 and taking the form $f(a,A,S) = \sum_{(u,t)\in S} g(a,\{u(a')\}_{a'\in A}))$, where the function g satisfies $g(a,\{\alpha u(a')\}_{a'\in A}) = \phi(\alpha)g(a,\{u(a')\}_{a'\in A})$ for all $\alpha \in \mathbb{R}$ and some invertible and odd $\phi : \mathbb{R} \to \mathbb{R}$.

[12] 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 (e.g., one could switch choices on the sets $\{a,b,c\}$, $\{a,b,d\}$, $\{a,b,e\}$, and $\{a,b\}$ to a), while ours counts three (not considering $\{a,b\}$ a violation).

This says the unit in which the preference intensity of different selves is measured does not affect rankings. This class includes utilitarianism as well as various menu-dependent variations. As previously noted, utilitarianism explains only rational choice behavior. This section shows that being able to explain *only* a limited set of behaviors is a nongeneric feature of aggregators in this class.

Consider the following model of reference-dependent aggregation in $\mathcal{F}^*$.

Example 6 - Simple reference dependence. The aggregate utility of an alternative a in a choice set A is $\sum_{(u,t)\in S}(u(a) - \text{mean } u(A))^\rho$, where ρ is an odd integer and mean $u(A)$ is a geometric or arithmetic mean over the set $\{u(a')\}_{a'\in A}$. This is a reference-dependent variation of the CRRA form, where the origin is shifted.

The reference dependence in Example 6 permits that model to rationalize a much wider array of behaviors than can utilitarianism. To understand why, let us first examine choice behavior over only three alternatives. There are three possible kinds of irrational choice functions defined over a three-element set. One possibility is transitive choice, where the second-best element (from the transitive ranking) is chosen from the triple; another is transitive choice, where the worst element is chosen from the triple; and the third is intransitive choice. Using the model in Example 6, it is easy to construct rationalizations for all three of these behaviors.

The first part of the following theorem shows that if a model of aggregation in $\mathcal{F}^*$ can rationalize the last two irrational behaviors over a triple of alternatives, then it can rationalize any choice function defined over any space of alternatives. The second part of the theorem shows that a generic aggregator in $\mathcal{F}^*$ (including Example 6) can rationalize any choice behavior with a uniform bound on the number of utility functions needed. To describe the metric for which genericity is defined, note that by scale invariance there is a natural bijection (simply by scaling the utility functions inputted) between (1) models in $\mathcal{F}^*$ applied to pairs and triples of elements, and (2) the set of pairs of operators $\Omega = \{O_1, O_2 \mid O_1 : \Delta_2 \to \mathbb{R}^2,\ O_2 : \Delta_3 \to \mathbb{R}^3\}$, where Δ_2, Δ_3 are the 2- and 3-dimensional simplices, respectively. The distance between two such pairs (O_1, O_2) and (O_1', O_2') is defined as $\max_{i=1,2} \sup_{x\in\mathbb{R}^i} |O_i(x) - O_i'(x)|$.

Theorem 1. *Let X be a finite grand set of alternatives. Then:*

(i) Take any model in $\mathcal{F}^*$ and any $x, y, z \in X$. If the model can rationalize both (1) intransitive choice over x, y, z and (2) transitive choice over x, y, z where the worst pairwise element is best in the triple $\{x, y, z\}$, then the model can rationalize any choice function c defined over X.

(ii) The set of models in $\mathcal{F}^*$ that can rationalize any choice function c using at most $1 + 5 \cdot \text{IIA}(c)$ utility functions is open and dense.

The proof of this theorem appears in the Appendix, and is discussed in the next section. Theorem 1 formalizes the sense in which only being able to explain rational choice behavior is fragile.

Once certain types of irrational behaviors can be explained over three alternatives, an additive and scale-invariant model can rationalize any choice behavior with sufficiently many selves. Moreover, the ability to explain any behavior is generic in this class, with at most five "good reasons" needed for every "mistake" made. Note that the result gives a lower bound on the set of behaviors a generic aggregator in $\mathcal{F}^*$ can rationalize, thereby providing a linear connection between the complexity of the observed behavior (as measured by the number of IIA violations) and the degree of freedom in the model (as measured by the number of utility functions). Given n utility functions, a generic aggregator in $\mathcal{F}^*$ can rationalize any choice function c, defined on any finite grand set of alternatives X, that has at most $\frac{n-1}{5}$ IIA violations. Thus, in spite of having a structured form, essentially any aggregator in $\mathcal{F}^*$ can rationalize any choice function with sufficiently many utility functions. In other words, a model of decision-making satisfying the above properties must put *a priori* restrictions on the number of utility functions in order to be falsifiable.

Given a model of aggregation and any triple of alternatives, it is very easy to check whether the model can rationalize the two irrational behaviors described in part (i) of Theorem 1. But the proof of Theorem 1 also reveals a simple sufficient condition for checking whether a model f is of the generic type in part (ii). It suffices to find a single self defined over a triple $\{x, y, z\}$ for which f "stretches" the utility differences over pairs,

$$\begin{aligned} &f(x,\{x,z\},s) - f(z,\{x,z\},s) \neq \\ &\quad f(x,\{x,y\},s) - f(y,\{x,y\},s) + f(y,\{y,z\},s) - f(y,\{y,z\},s), \end{aligned}$$

and for which f's evaluation of alternatives in the triple is not fixed by the pairwise rankings,

$$\begin{aligned} &f(x,\{x,y,z\},s) - f(y,\{x,y,z\},s) + f(x,\{x,y,z\},s) - f(z,\{x,y,z\},s) \neq \\ &\quad f(x,\{x,y\},s) - f(y,\{x,y\},s) + f(x,\{x,z\},s) - f(z,\{x,z\},s). \end{aligned}$$

For example, defining the above self using the utility function $u(y) = 4 > u(z) = 2 > u(x) = 1$ shows that the model in Example 6 using an arithmetic mean is in the generic class. By contrast, utilitarianism and generalizations of the form $f(a, A, s) = u(a) + h(A)$, where the choice set cannot change intensity of preference within a set, fail the sufficient condition (and, in fact, explain only rational choice). The proof shows that the sufficient condition is satisfied generically. Nonetheless, it is not necessary – even aggregators that fail to satisfy the condition may be able to rationalize all choice behaviors. As seen from our upcoming results, the model of Example 2 using linear Φ can rationalize any behavior with five utility functions per IIA violation, but fails the sufficient condition.

5 A rationalization theorem and procedure

We begin with an illustrative example before presenting our main result. Recall the model in Example 2, where the aggregate utility of an alternative $a \in A$ is

$$f(a, A, S) = \sum_{(u,t)\in S} \Phi(\max_{b\in A} u(b) - \min_{b\in A} u(b))u(a)$$

for some monotonic function Φ. Let us suppose Φ is increasing, and examine how this aggregator behaves on an arbitrary three-element set of alternatives $\{a, b, c\}$. In particular, define a collection S of five selves having the following five utility functions defined on $\{a, b, c\}$ (in each column, the alternative on the left receives the utility number to its right):

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 a is chosen from $\{a, b\}$. Indeed, $f(a, \{a, b\}, S) = 4\Phi(2) + \Phi(1)$ and $f(b, \{a, b\}, U) = 2\Phi(2) + 3\Phi(1)$. Hence $f(a, \{a, b\}, S) > f(b, \{a, b\}, S)$ since $\Phi(2) > \Phi(1)$. In any other choice set, all alternatives have the same aggregate utility:

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

That is, under the collection of selves S, alternative a receives strictly higher aggregate utility than b in the choice set $\{a, b\}$, and there is complete indifference in all other choice problems. We will call such a collection S defined on a three-alternative set $\{a, b, c\}$ a *triple-basis* for this aggregator f. Triple-bases can serve as building blocks for rationalizations of choice functions on arbitrary spaces of alternatives. To illustrate, take $X = \{x_1, x_2, \ldots, x_n\}$ and define the choice function $c(\cdot)$ as selecting the alternative with the smallest index in every choice set, with the exception that $c(\{x_i, x_j\}) = x_j$ for one pair $i < j$. This choice function has one IIA violation, corresponding to the set $\{x_i, x_j\}$. Using the triple basis above, we construct a collection of five selves $S^{\{x_i,x_j\}}$ having the following utility functions over X:

u_1		u_2		u_3		u_4		u_5	
x_i	2	x_i	2	$X \setminus \{x_i, x_j\}$	2	$X \setminus \{x_i\}$	2	x_j	2
$X \setminus \{x_i, x_j\}$	1	x_j	1	x_i	1	x_i	0	$X \setminus \{x_j\}$	0
x_j	0	$X \setminus \{x_i, x_j\}$	0	x_j	0				

This is constructed by letting the choice from $\{x_i, x_j\}$, which is x_j, play the role of a in the triple-basis; letting the unchosen alternatives in $\{x_i, x_j\}$, which is only x_i, play the role of b; and letting the alternatives outside $\{x_i, x_j\}$, which consists of $X \setminus \{x_i, x_j\}$, play the role of c.

Using $S^{\{x_i,x_j\}}$, how does f evaluate the alternatives in each choice set $A \subseteq X$? Since any $x \in X \setminus \{x_i, x_j\}$ has the same utility as c in the calculations above, it is easy to see $f(\cdot, A, S^{\{x_i,x_j\}})$ is constant for any set $A \neq \{x_i, x_j\}$. Since x_j plays the role of a and x_i plays the role of b, the previous calculations imply that $f(x_j, \{x_i, x_j\}, S^{\{x_i,x_j\}}) > f(x_i, \{x_i, x_j\}, S^{\{x_i,x_j\}})$. Thus, the utility functions in $S^{\{x_i,x_j\}}$ rationalize the choice from $\{x_i, x_j\}$ and have no impact on other choice sets.

Since the collection $S^{\{x_i,x_j\}}$ has implications only for the IIA violation $\{x_i, x_j\}$, one needs an additional self to rationalize the remaining "rational" choices. We construct a final self s^* whose utility function u^* has sufficiently small range to not overturn any strict preferences induced from $S^{\{x_i,x_j\}}$, and which has the ranking $u^*(x_1) > u^*(x_2) > \cdots > u^*(x_n)$ derived from standard revealed preference. By construction, the selves $\langle s^*, S^{\{x_i,x_j\}}\rangle$ rationalize $c(\cdot)$.

5.1 Rationalizability result

Observe that the triple basis S given above would still be a triple-basis for the generalized additive difference model if we were to scale all the utilities by a common constant. Loosely speaking, this means that for any δ, the collection of selves S rationalizes being indifferent among all alternatives in subsets of $\{a, b, c\}$ except for having a δ-amount of strict preference within one pair. This is a property we term *triple-solvability*, and is formally defined below for any model of aggregation.

Definition 4. *Given a triple $\{a, b, c\}$ and model (f, T), the collection of selves $S \in \mathcal{S}(\{a, b, c, \}, T)$ is a triple-basis if $f(a, \{a, b\}, S) > f(b, \{a, b\}, S)$ and $f(\cdot, A, S)$ is constant for all other $A \subseteq \{a, b, c\}$. The model (f, T) is triple-solvable with k utility functions if for every $\delta > 0$, there is a triple-basis $S \in \mathcal{S}^k(\{a, b, c\}, T)$ with $\max_{a,b\in A, A\subseteq\{a,b,c\}, s\in S} |f(a, A, s) - f(b, A, s)| < \delta$.*

Given a model, it is easy to check for the existence of a triple-basis. Indeed, triple bases can be found for the models featured earlier.[13] 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

[13] Solvability of the simple reference dependence model will follow from the sufficient condition it satisfies. For the case of the contextual concavity model of Kivetz et al. (2004), the following is a triple basis for any $\rho \neq 1$: $u_1(a) = 4, u_1(b) = 3, u_1(c) = 1$, $u_2(a) = 3, u_2(b) = 1, u_2(c) = 2$, $u_3(a) = 3, u_3(b) = 4, u_3(c) = 1$, $u_4(a) = 1, u_4(b) = u_4(c) = 3$, $u_5(a) = 2, u_5(b) = 1, u_5(c) = 3$, $u_6(a) = 1, u_6(b) = 2, u_6(c) = 4$. For the case of loss aversion with kinked linear ℓ and parameter 2, the following is a triple basis (there is some rounding error): $u_1(a) = -2.112, u_1(b) = -1.275, u_1(c) = 7.225$, $u_2(a) = 0, u_2(b) = 1.445, u_2(c) = 1$, $u_3(a) = 6, u_3(b) = 7.225, u_3(c) = 4$, $u_4(a) = -4.766, u_4(b) = -2.938, u_4(c) = 0$, $u_5(a) = 5, u_5(b) = -5.981, u_5(c) = 2.814$. For bargaining with endogenous disagreement point, the following is a triple basis (there is some rounding error): $u_1(a) = 2.847, u_1(b) = 1, u_1(c) = 7.634$, $u_2(a) = 0, u_2(b) = 4.288, u_2(c) = 1$, $u_3(a) = 6, u_3(b) = -.129, u_3(c) = 4$, $u_4(a) = -4.651, u_4(b) = -.949, u_4(c) = 0$, $u_5(a) = 5, u_5(b) = -1.619, u_5(c) = -15.8$.

can be scaled as needed. More generally, it is easy to see from our construction that it suffices for there to be triple-bases using only $|X| - 2$ δ's, where each is smaller than the amount of strict preference under the previous δ's. It turns out that triple solvability holds broadly among the class of models featured here, and in fact models in the class $\mathcal{F}^*$ generically satisfy this property. The fact that these examples illustrate various models of multi-self decision-making proposed in the literature suggests that this property, which can be checked simply by looking at choice behavior on three-element sets, holds broadly. As our next result shows, this behavioral property has strong implications for the explanatory power of a model.

Theorem 2. *Suppose the model $(f, T) \in \mathcal{F}$ is triple-solvable with k_f selves. Then, for any choice function c, defined on any finite grand set of alternatives X, no more than $1 + k_f \cdot \text{IIA}(c)$ selves are needed to rationalize c.*

We sketch below the proof of Theorem 2, describing our general rationalization method and its connection to our definition of an IIA violation. Note that an alternative statement of the result is as follows: using n selves, the model (f, T) can rationalize any choice function c, defined on any finite grand set of alternatives X, which has at most $\frac{n-1}{k_f}$ IIA violations. Hence, the result also gives a lower bound on the set of rationalizable behaviors for a fixed number of selves, providing a linear connection between the number of IIA violations and the degree of freedom in the model (as measured by the number of selves). In Supplementary Appendix B, we apply this result to understand a generalized Strotzian model: a DM chooses a menu knowing her choice from the menu is a compromise among multiple motivations. Behavior that is interpreted differently by the literature on choice over menus can arise from "anticipated" IIA violations.

Note that for each model (f, T), the proportionality constant k_f is independent of the size of the alternative space X, and can be calculated using any triple of alternatives (it is simply the number of selves in a triple basis). This means that the number of selves that are sufficient to rationalize a choice function on the alternative space X does not increase if the choice function is extended to a larger alternative space $\hat{X}$ in a manner such that no additional IIA violations are created. This formalizes the sense in which the size of the rationalization depends directly on the complexity of the behavior and not the size of the alternative space; the size of the alternative space matters only in the sense that it bounds the number of IIA violations that are possible.

Sketch of proof: a universal rationalization method. Suppose f is triple-solvable with k_f utility functions. Given an arbitrary X and any choice function c defined on X, 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 causing an IIA violation, the construction creates a collection of selves S^A whose utility functions, defined on X, correspond to those of a triple basis: $c(A)$ plays the role of the preferred element a in $\{a, b\}$, $A \setminus \{c(A)\}$ plays the role of the unchosen

element b in this pair, and $X \setminus A$ plays the role of the third element c. The properties P1 and P5 (neutrality and profile equivalence) imply that the selves S^A behave similarly to the triple basis:

1. Under the model, the selves S^A imply an aggregate strict preference for $c(A)$ in every subset of A in which it is contained; and

2. The selves S^A are completely indifferent between all options for all other choice sets; that is, sets not containing $c(A)$ or sets containing some element of $X \setminus A$.

Remember that a set A causes an IIA violation if there is a superset B such that $c(B) \in A \setminus \{c(A)\}$, and there is no set in between A and B (in terms of containment) which has an IIA implication for A. When we rationalize the choice from A using the selves S^A, there is a "trickle down" effect: following IIA, those selves continue to prefer $c(A)$ in subsets of A. Another IIA violation may occur for some subset of $A' \subset A$, where $c(A)$ is available but not chosen. In the recursive procedure, the selves $S^{A'}$ corresponding to the IIA violation in the smaller set A' are constructed first. Upon reaching the set A, the triple-basis used to generate S^A must be indifferent enough over the alternatives so that the trickle-down effect of S^A does not overturn the strict preference of $S^{A'}$. This is possible by P4 (continuity to near-indifferent additions). Since the selves S^A only induce strict aggregate preference in subsets of A which contain $c(A)$, P3 (reinforcement) implies they do not affect the aggregate preference over alternatives in other sets. In particular, the selves S^A do not interfere with selves constructed for any other IIA violations. Similarly, the selves $S^{A'}$ do not interfere with the choice from larger sets, such as A or X. Of course, since the selves constructed for IIA violations have no implications for larger sets (such as X), there must be some self which accounts for those remaining choices as well. The construction is completed by creating an extra self s^* which, using P4, is indifferent enough never to overturn any strict preferences from selves associated with IIA violations. Using P2 (consistency), this self is constructed via standard revealed preference, by allocating the highest utility to $c(X)$, the next highest utility to $X \setminus \{c(X)\}$, etc.

To summarize, since there are IIA(c) violations in the observed choice behavior, and each triple basis has k_f selves, this procedure thus generates $1 + k_f \cdot \text{IIA}(c)$ selves. This collection of selves rationalizes the observed choice behavior $c(\cdot)$ under the model. As shown more formally in the appendix, the selves generated for a set which causes an IIA violation ensure that the choice from that set is indeed picked under the model. Furthermore, their *only* other effect is to ensure that particular alternative continues to be picked in subsets in which it is contained – unless there is a subset which causes another IIA violation. In that case, selves are constructed which overturn any aggregate preferences coming from larger sets. All "rational" choices which are not otherwise covered by the selves corresponding to IIA violations are implied by the final, rational self s^*.[14] ∎

[14] While our goal is to show that any behavior can be rationalized with sufficiently many selves, it is easy see that the above construction also goes through using a refinement of our definition of IIA violations, showing even more behaviors are rationalizable with a given number of selves. This refinement says a set A (which would have otherwise been a violation) does not count as one if all the minimal supersets of A that cause an IIA violation also choose $c(A)$.

It is easy to see that the proposed rationalization procedure can be modified to generate rationalizations of choice correspondences, by extending our definition of IIA violations for choice functions to count both violations of Sen's α and Sen's β (axioms that, when taken together, are equivalent to rational choice behavior for correspondences).

Theorem 1, for scale-invariant and type-independent aggregators, is proved in five steps. The first is knowing that if f is triple-solvable with k selves, we can rationalize any choice function c with $1+k\cdot\text{IIA}(c)$ selves. This is simply Theorem 2. The next step is showing that if a certain matrix – constructed by permuting possible aggregate utility differences given various rankings of three alternatives a, b and c – has a nonzero determinant, then the aggregator f is triple-solvable. Next, we prove part (i) by showing that if the two types of irrational behaviors can be explained, then the above matrix has nonzero determinant. To prove part (ii), we first show that if the sufficient condition described after Theorem 1 is satisfied, then the above matrix has nonzero determinant *and* the aggregator is triple-solvable with five utility functions. Finally, we prove the sufficient condition is generically satisfied. For intuition on why the bound is five, notice that checking whether a collection constitutes a triple basis requires checking five aggregate utility differences: the aggregate utility difference between any two pairs of alternatives within the set $\{a, b, c\}$, and the aggregate utility difference between the alternatives within each of the three pairs $\{a, b\}$, $\{b, c\}$, and $\{a, c\}$. It turns out that a generic model in $\mathcal{F}^*$ "stretches" utility differences in a nonlinear, menu-dependent fashion, and that under scale-invariance, having five selves provides enough degrees of independence to ensure that a triple basis can be constructed.

6 Discussion

This paper studies a framework that encompasses many multi-self models proposed in the literature. Our results have implications in both interpersonal and intrapersonal decision-making, calling attention to the importance of collecting reliable information on the number of selves (motivations, in the interpersonal context) participating in the decision. We identify a class of models for which it is important to impose a priori restrictions on the number of selves in order to ensure falsifiability; outside this class of models, one can find examples where such restrictions are not needed. To our knowledge, the models treated here have not been characterized from a choice-theoretic perspective. Indeed, the fact that all selves contribute to aggregate utility on every choice set can make it difficult to construct a rationalization of a choice function, or ascertain where a rationalization exists. The proof of our result provides a universal procedure for constructing such a rationalization.

Appendix

Proof of Theorem 2

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 $I_1 = \{A_1^1, ..., A_{i_1}^1\}$ 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 $I_j = \{A_1^1, ..., A_{i_{j+1}}^1\}$ 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 $\bigcup_{l=1}^{j-1} I_l$ 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 $\bar{S}^1 = \langle \bar{s}_1^1 \dots, \bar{s}_k^1 \rangle$ that is a triple basis over $\{a, b, c\}$ (existence follows from triple-solvability). For every $A \subset I_1$, construct now the following collection $S^A = \langle s_1^A, \dots s_k^A \rangle$ where each self s_i^A has type $\bar{t}_i$ and utility function u_i^A defined over X by

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

Suppose now that S^A is defined for every $A \in \bigcup_{k=1}^{j} I_k$ for some $j \geq 1$. Let S_k be the collection of selves $S_k = \langle S^{A_1^k}, ..., S^{A_{i_k}^k} \rangle$, for $k = 1, ..., j$. Let $\widehat{S}_j = \langle S_1, ..., S_j \rangle$. By P4, there exists $\delta > 0$ such that for any δ-indifferent collection of k selves S', $f(a, A, \widehat{S}_j) > f(b, A, \widehat{S}_j)$ implies $f(a, A, \langle \widehat{S}_j, S' \rangle) > f(b, A, \langle \widehat{U}_j, U' \rangle)$. Then by P3 and P6, we know

$$\begin{aligned} &f(a, A, \langle \widehat{S}_j, \widetilde{S}_1, ..., \widetilde{S}_m \rangle) > f(b, A, \langle \widehat{S}_j, \widetilde{S}_1, ..., \widetilde{S}_m \rangle) \text{ implies} \\ &f(a, A, \langle \widehat{S}_j, \widetilde{S}_1, ..., \widetilde{S}_m, S' \rangle) > f(b, A, \rangle \widehat{S}_j, \widetilde{S}_1, ..., \widetilde{S}_m, S' \rangle) \end{aligned}$$

for any (exactly) indifferent collections of selves $\widetilde{S}_1, ..., \widetilde{S}_m$. Let now $I_{j+1} = \{A_1^1, ..., A_{i_{j+1}}^1\}$ 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 δ-indifferent collection of k selves $\bar{S}^{j+1} = \langle \bar{s}_1^{j+1}, \dots \bar{s}_k^{j+1} \rangle$ that is a triple basis over $\{a, b, c\}$. For every $A \in I_{j+1}$, construct the collection of selves $S^A = \langle s_1^A, \dots s_k^A \rangle$ where each self s_i^A

has type $\bar{t}_i^{j+1}$ and the utility function u_i^A over X defined by:

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

for every $i = 1, ..., k$. Let S_{j+1} be the collection $\langle S_j, S^{A_1^1}, ..., S^{A_{i_{j+1}}^1} \rangle$. 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 δ_{j^*}-indifferent self s^*, $f(a, A, S_{j^*}) > f(b, A, S_{j^*})$ implies $f(a, A, \langle S_{j^*}, s^* \rangle) > f(b, A, \langle U_{j^*}, s^* \rangle)$. Finally, construct one more self s^* in the following way. Let $a_1 = c(X)$ and $a_k = c(X \setminus \{a_1, a_2, \dots a_{k-1}\})$ for $2 \leq k \leq n$. Fix some $t^* \in T$ and let $s^* = (t^*, u^*)$, where we construct $u^* : X \to \mathbb{R}$ such that $u^*(a_1) > u^*(a_2) > \cdots > u^*(a_n)$ and u^* is δ_{j^*}-indifferent. We show that the collection $S_c \equiv \langle S_{j^*}, s^* \rangle$ rationalizes c under aggregator f.

Observation 1. *For any set A which is an IIA violation, by P1, P5, and construction of S^A, $f(a, B, S^A) = f(b, B, S^A)\ \forall\ B$ and $a, b \in B$ such that $B \setminus A \neq \emptyset$ or $c(A) \notin B$, and $f(c(A), B, S^A) > f(b, B, S^A) = f(b', B, S^A)\ \forall\ b, b' \in B \setminus \{c(A)\}$ 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 1, $f(a, X, S^A) = f(b, X, S^A)$ for every $a, b \in X$ and every A with which there is an IIA violation associated. Moreover, $f(c(X), X, s^*) > f(a, X, s^*)\ \forall\ a \in X \setminus \{c(X)\}$ by P2. Then repeated application of P3 implies $f(c(X), X, S_c) > f(a, X, S_c)\ \forall\ a \in X \setminus \{c(X)\}$. Next, consider any $A \subsetneq X$ which causes an IIA violation. Suppose $A \in I_j$. Observation 1 implies that for any $B \in (\bigcup_{l=1}^{j} I_l) \setminus A$, $f(a, A, S^B) = f(a', A, S^B)\ \forall\ a, a' \in A$, and $f(c(A), A, S^A) > f(a, A, S^A)\ \forall\, a \in A$. Then repeated implication of P3 implies $f(c(A), A, S_j) > f(a, A, S_j)\ \forall\ a \in A$. By construction then $f(c(A), A, S_c) > f(a, A, S_c)\ \forall\ a \in A$. Finally, consider a set A that does not cause an IIA violation. There are several cases.

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

Case 2: There is a unique $a \in A$ such that for some $B \supset A$, $c(B) = a$. First, note that $a = c(A)$ is necessary, otherwise A would have caused a 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, $u^*(c(A)) > u^*(a)$ for all $a \in A \setminus \{c(A)\}$, and $f(a, A, S^B) = f(a', A, S^B)\ \forall\ a, a' \in A$ and B with which an IIA violation is associated. Hence, $f(c(A), A, S_c) > f(a, A, S_c)$ for all $a \in A$.
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 1, for any $A \in \bigcup_{l=1}^{j} I_l$ either $f(c(A), A, S^B) > f(a, A, S^B)$ for all $a \in A$, or $f(a, A, S^B) = f(a', A, S^B)$ for all $a, a' \in A$. But then repeated application of P3 implies that $f(c(A), A, S_j) > f(a, A, S_j)$ for all $a \in A$. By construction, $f(c(A), A, S_c) > f(a, A, S_c)$ for all $a \in A$.

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 $c(A)$, otherwise A would have caused an IIA violation. Let $\{b_i\}_i$ be the set of elements other than $c(A)$ 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 $B_i \supset B_m$ 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 $c(A'_j) = c(A)$. 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 $c(A'') \in A'_j$, then $c(A'') \neq c(A)$, by maximality of A'_j. 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, we added selves to ensure choice $c(A)$ for every A'_j. Thus a should be the choice from A unless selves were added for some B' between some A'_j and A for which $c(B') \in A \setminus \{a\}$. This is impossible by minimality of the B_j's. ∎

Proof of Theorem 1

Theorem 1 follows from Theorem 2 and the following three lemmas. Let $X = \{a, b, c\}$ and take any $f \in \mathcal{F}^*$. For compactness, we use the notation $x_1 = f(a, \{a, b, c\}, S) - f(b, \{a, b, c\}, S)$, $x_2 = f(b, \{a, b, c\}, S) - f(c, \{a, b, c\}, S)$, $x_3 = f(a, \{a, c\}, S) - f(c, \{a, c\}, S)$, $x_4 = f(b, \{b, c\}, S) - f(c, \{b, c\}, S)$, and $x_5 = f(a, \{a, b\}, S) - f(b, \{a, b\}, S)$.

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

Proof. Consider the following table.

$1: S$	$2: (bc)(a)$	$3: (ab)(c)$	$4: (abc)$	$5: (acb)$	$6: a \sim b \succ c$	$7: a \succ b \sim c$
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

Column 1 lists aggregate utility values under S. By neutrality, if we can generate column 1, we can also generate the 2nd column using the permutation $(bc)(a)$ over alternatives, the 3rd column

using the permutation $(ab)(c)$ over alternatives, etc. We generate columns 6 and 7 using profile equivalence to evaluate $f \circ u$ and $f \circ u'$ for the utility functions u and u' given by those headers. Determinants of three possible 5×5 matrices, each composed of five of the columns above, are:

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

To prove the result, it suffices to show that there exists S such that defining $x_1, x_2, \ldots, x_5$ 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 utilities of selves corresponding to the i-th column are multiplied by c_i. The resulting collection is 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 selves S defined over $\{a, b, c\}$ such that $x_3 \neq x_4 + x_5$ and which rationalize under f the choice behavior 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 $x_1 + 2x_2 \neq x_3 + x_4$.*[15]

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 LHS$< 0$ and RHS> 0.
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 S such that f rationalizes this behavior using the selves S, then observe that $x_1 + 2x_2$ is negative. Hence $x_1 + 2x_2 \neq x_3 + x_4$ because RHS> 0. ∎

Say that $f \in \mathcal{F}^*$ is *non-degenerate* if for some utility function u on $X = \{a, b, c\}$, we have $x_3 \neq x_4 + x_5$ and $2x_1 + x2 \neq x_3 + x_5$ using the collection S consisting of one self with utility u. We formally establish that for any fixed scaling function $\phi(\alpha)$ the property that an additive, neutral and scale-invariant aggregator $f \in \mathcal{F}^*$ is nondegenerate holds generically. In order to define a topology on $\mathcal{F}^*$, we transform the latter set of aggregators to a convenient representation. Note that for a fixed scaling function, specifying the aggregated utilities of n alternatives for selves in the n-dimensional simplex determines the aggregated utilities of n alternatives for all possible selves

[15] 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 S such that $f \circ S$ rationalizes this behavior, then $x_3, x_4, x_5 > 0$ and $x_1 \leq 0$, $x_2 > 0$. Observe that $2x_1 + x_2 < 0$. Therefore, $2x_1 + x_2 \neq x_3 + x_5 > 0$.

over n alternatives, since any self is a scalar multiple of exactly one self from the simplex. Hence, with respect to a grand set of alternatives with three elements, there is a natural bijection β between additive and scale-invariant aggregators, and the set of pairs of operators $\Omega = (O_1, O_2 | O_1 : \Delta_2 \to \mathbb{R}^2; O_2 : \Delta_3 \to \mathbb{R}^3)$, where O_1 determines how a self's utilities get aggregated in pairs, and O_2 determines how a self's utilities get aggregated in the triple. Define metric d on Ω such that the distance between (O_1, O_2) and (O_1', O_2') is defined as $\max_{i=1,2} \sup_{x \in \mathbb{R}^i} |O_i(x) - O_i'(x)|$.

Lemma 3. *Given the topology induced by d, the pairs of operators in Ω that are associated with non-degenerate aggregators in $\mathcal{F}^*$ is open and dense relative to Ω.*

Proof. For any utility function v and $f \in \mathcal{F}^*$, let s be a self with utility v and define

$$\begin{aligned}
\Gamma_1^l(f, v) &= f(a, \{a, c\}, s) - f(c, \{a, c\}, s), \\
\Gamma_1^r(f, v) &= f(a, \{a, b\}, s) - f(b, \{a, b\}, v) + f(b, \{b, c\}, s) - f(c, \{b, c\}, s), \\
\Gamma_2^l(f, v) &= f(a, \{a, b, c\}, s) - f(b, \{a, b, c\}, s) + f(a, \{a, b, c\}, s) - f(c, \{a, b, c\}, s), \\
\Gamma_2^r(f, v) &= f(a, \{a, b\}, s) - f(b, \{a, b\}, s) + f(a, \{a, c\}, s) - f(c, \{a, c\}, s),
\end{aligned}$$

1. Openness. Suppose f is nondegenerate and let $u : \{a, b, c\} \to \mathbb{R}$ satisfy the nondegeneracy requirement. Note that u cannot be fully indifferent. Let $\varepsilon_i = \Gamma_i^l(f, u) - \Gamma_i^r(f, u)$ for $i \in \{1, 2\}$, and let $\varepsilon = \max(|\varepsilon_1|, |\varepsilon_2|)$. Next, for every $i, j \in \{a, b, c\}$ such that $i \neq j$, let α^{ij} be such that $\alpha^{ij}(u(i), u(j)) \in \Delta^2$. Note that the terms α^{ij} are uniquely defined. Similarly, let α^{abc} be such that $\alpha^{abc}(u(a), u(b), u(c)) \in \Delta^3$. Let $\alpha = \max(|\alpha^{ab}|, |\alpha^{ac}|, |\alpha^{bc}|, |\alpha^{abc}|)$. Since s is not an indifferent self, $\alpha > 0$. Then for $\delta < \frac{\varepsilon}{8\alpha}$ it holds that $\Gamma_i^l(f', u) \neq \Gamma_i^r(f', u)$ for $i \in \{1, 2\}$ for every f' such that $|\beta(f) - \beta(f')| < \delta$, since each term given f' in the above inequalities can differ from the corresponding term given f by at most $\frac{\varepsilon}{8}$.

2. Denseness. Let $\delta > 0$. Consider a self s with utility $u \in \Delta_3$ over $\{a, b, c\}$ such that $u(a) > u(b) > u(c)$. For every $i, j \in \{a, b, c\}$ such that $i \neq j$, let α^{ij} satisfy $\alpha^{ij}(u(i), u(j)) \in \Delta^2$. Let $\alpha = \max(|\alpha^{ab}|, |\alpha^{ac}|, |\alpha^{bc}|)$. If nondegeneracy holds, there is trivially a point in the δ-neighborhood of $\beta(f)$ corresponding to a nondegenerate aggregator. Otherwise let $\varepsilon \in (0, \frac{\delta}{\alpha})$ be such that $\varepsilon \neq |\Gamma_i^l(f, u) - \Gamma_i^r(f, u)|$ for $i \in \{1, 2\}$. Take any $f' \in \mathcal{F}^*$ for which (i) for triples, f' is equivalent to f; and (ii) for a pair $\{x, y\}$, given any utility function v over $\{x, y\}$ for which $v(x) \geq v(y)$, $f'(x, \{x, y\}, v) = f(x, \{x, y\}, v)$ and $f'(y, \{x, y\}, v) = f(y, \{x, y\}, v)$ if $v(x) - v(y) < u(a) - u(c)$, but $f'(x, \{x, y\}, v) = f(x, \{x, y\}, v) + \varepsilon$ and $f'(y, \{x, y\}, v) = f(y, \{x, y\}, v)$ if $v(x) - v(y) \geq u(a) - u(c)$. Thus, with respect to selves for which the utility difference between elements of the pair is at least $u(a) - u(c)$, aggregate utility is $\varepsilon > 0$ higher than what f yields for the preferred alternative (but is the same for other alternative) - otherwise f' is equivalent to f. By construction, $|\beta(f') - \beta(f)| < \delta$. Also, $\Gamma_1^l(f', v) = \Gamma_1^l(f, v) + \varepsilon$, $\Gamma_1^r(f', v) = \Gamma_1^r(f, v)$, $\Gamma_2^l(f', v) = \Gamma_2^l(f, v)$, and $\Gamma_2^r(f', v) = \Gamma_2^r(f, v) + \varepsilon$. Then $\varepsilon \neq |\Gamma_i^l(f, v) - \Gamma_i^r(f, v)|$ for $i \in \{1, 2\}$ implies that $\Gamma_i^l(f', v) \neq \Gamma_i^r(f', v)$ for $i \in \{1, 2\}$. Hence, f' is non-degenerate. ∎

References

Apps, Patricia and Ray Rees, "Taxation and the Household," *Journal of Public Economics*, 1988, *35*, 355–369.

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.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen, "The Collective Model of Household Consumption: A Nonparametric Characterization," *Econometrica*, 2007, *75*, 553–574.

___ , ___ , and ___ , "Opening the Black Box of Intra-Household Decision-Making: Theory and Non-Parametric Empirical Tests of General Collective Consumption Models," *Journal of Political Economy*, 2009, *117*, 1074–1104.

___ , ___ , and ___ , "The Revealed Preference Approach to Collective Consumption Behavior: Testing and Sharing Rule Recovery," *Review of Economic Studies*, 2011, *78*, 176–198.

Cherepanov, Vadim, Tim Feddersen, and Alvaro Sandroni, "Rationalization," *Theoretical Economics*, forthcoming.

Chiappori, Pierre-André, "Rational Household Labor Supply," *Econometrica*, 1988, pp. 63–89.

___ and Ivar Ekeland, "The Microeconomics of Group Behavior: General Characterization," *Journal of Economic Theory*, 2006, *130*, 1–26.

Conley, John, Richard McLean, and Simon Wilkie, "Reference Functions and Possibility Theorems for Cardinal Social Choice Problems," *Social Choice and Welfare*, 1997, *14*, 65–78.

de Clippel, Geoffroy and Kfir Eliaz, "Reason Based Choice: a Bargaining Rationale for the Attraction and Compromise Effects," *Theoretical Economics*, 2012, *7*, 125–162.

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.

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*, 14031436.

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*, 24812488.

Kaneko, Mamoru and Kenjiro Nakamura, "The Nash Social Welfare Function," *Econometrica*, 1979, *47*, 423–435.

Karni, Edi, "Impartiality: Definition and Representation," *Econometrica*, 1998, *66*, 1405–1415.

Kőszegi, Botond and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics, forthcoming*, 2012.

Kivetz, Ran, Oded Netzer, and V. Srinivasan, "Alternative Models for Capturing the Compromise Effect," *Journal of Marketing Research*, 2004, *41*, 237–257.

Kreps, David, "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 1979.

Manzini, Paola and Marco Mariotti, "Sequentially Rationalizable Choice," *American Economic Review*, 2007, *97*, 1824–1839.

May, Kenneth O., "Intransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica*, 1954, pp. 1–13.

Orhun, Yesim, "Optimal Product Line Design When Consumers Exhibit Choice Set Dependent Preferences," *Marketing Science*, 2009, *28*, 868–886.

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.

Supplementary Appendices, Not for Publication

This document contains supplementary appendices to "Rationalizing Choice with Multi-Self Models" by Ambrus and Rozen. The main paper is referenced throughout as AR.

A Examples rationalizing common choice procedures

Example 1 (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.

$$f(a, A, X, S) = \prod_{(u,t)\in S} (u(a) + \max_{a'\in X} u(a') - \underset{a'\in A}{med}\, u(a')),$$

where $\underset{a'\in A}{med}\, u(a')$ 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. We claim that with this aggregator, two selves can be used to rationalize the median procedure. Let $a_1, a_2, ..., a_N$ 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 $i \in \{1, ..., N\}$. 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 $u_1(a) + \max_{a'\in X} u_1(a') - \underset{a'\in A}{med}\, u_1(a')$ and $u_2(a) + \max_{a'\in X} u_2(a') - \underset{a'\in A}{med}\, u_2(a')$ is constant across all elements of X, and the aggregated utility is 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.

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 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.

Another simple procedure Kalai et al. (2002) study is Sen (1993)'s second-best procedure.

Example 2 (Choosing the second best)**.** *Consider the following procedure: there is some strict ordering $\succ$ defined over elements of X, and the DM always chooses the second best element of any choice set, according to $\succ$. We will show that there is an aggregator that can rationalize the choice function given by the above procedure no matter how large X is, using only two selves. For any self u on X, and any $A \subset X$, let $l(u, A)$ be the lowest utility level attainable from A according to u. Moreover, let $g : X \times P(X) \times \mathcal{X} \times R^X \times T \to \mathbb{R}$ be such that*

$$g(a, A, X, s) = \begin{cases} u(a) - \max\limits_{b \in X} u(b) & \text{if } u(a) = l(u, A) \\ u(a) & \text{otherwise.} \end{cases}$$

That is, g penalizes the worst elements of a given choice set, by an amount that corresponds to the best attainable utility in X. Define now the following aggregator: for any collection of selves with utility functions $u_1, \ldots, u_n$ defined over X, let $f(a, A, X, S) = \sum\limits_{i=1}^{n} g(a, A, X, u_i)$. That is, f is a utilitarian aggregation, with large disutility associated with alternatives that are worst for some selves in the choice set. We claim that the following two selves rationalize the second-best procedure with f. Let $a_1, a_2, ..., a_N$ stand for the increasing ordering of alternatives in X according to $\succ$, and define $u_1(a_j) = j$ and $u_2(a_j) = N + \frac{N+1-j}{2N}$ for all $j \in \{1, ..., N\}$. Note that the incremental utilities of u_1 when choosing a higher $\succ$-ranked element are larger than the incremental disutilities of u_2. Hence this self determines the preference ordering implied by the aggregated utility, with the exception of the choice between the best alternative and the second-best alternative for u_1 in the choice set. This is because the best alternative for u_1 is the worst for u_2, and the extra disutility associated with this worst choice for u_2 overcomes the incremental utility for u_1. This rationalization has the simple interpretation of a conflict between a greedy self and an altruistic self.

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 either of the above procedures goes to infinity. Kalai et al. (2002) also discuss the idea that when multiple rationalizations are possible, one with the minimal number of selves is most appealing. While dictator-type aggregators do not provide an intuitively appealing explanation for the median procedure, aggregators in our framework can rationalize the above procedures in simple and intuitive ways. Note that the aggregators and selves in these examples together rationalize very specific

types of behavior. A given aggregator might act differently on different selves. For example, if the two selves did not have exactly opposing preferences in the example rationalizing the median procedure, the aggregator might not choose a centrally located alternative in every choice set. Hence AR studies the *set* of behaviors that an aggregator can rationalize (with different selves).

B 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 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.[16] 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.

Definition 5. *The DM's preference over menus $\succeq$ has a generalized Strotzian representation if there exists a collection of selves $S \in \mathcal{S}(X,T)$, an aggregator f, and a utility function $W : X \to \mathbb{R}$ on alternatives such that $\succeq$ is represented by the utility function $V : P(X) \to \mathbb{R}$ on sets, defined by*

$$V(A) = W\Big(\arg\max_{a\in A} f(a, A, S)\Big).$$

The generalized Strotzian representation has a straightforward interpretation: the DM picks the set for which the element foreseen to be chosen yields the greatest current utility.[17] Moreover, the DM expects to choose from the menu while subject to multiple motivations. 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 $\{a\}_{a\in X}$.

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

[16] We thank Eddie Dekel for suggesting the Strotzian interpretation of the representation.

[17] This relates to the separation of decision utility and experienced utility proposed by Kahneman, Wakker and Sarin (1997).

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)).

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

Corollary 3. *$\succeq$ satisfies Axioms 1-3 if and only if $\succeq$ has a generalized Strotzian representation using a collection of selves S 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*

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

the number of selves in the representation is no larger than $1 + k \cdot IIA(c_{\succeq})$.[18]

Proof. To prove this result, 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 then use the result of Theorem 2 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 the above:

Observation 2. *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 unforseen contingencies grows when there is additional desire for flexibility or self-control. IIA violations in anticipated choice are

[18] 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.

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.

C Approximate triple-solvability

While triple solvability is a property that is broadly satisfied, it can be seen from our construction that our rationalizability theorem would still hold under a weaker condition. It suffices that there exists a collection which is arbitrarily close to being indifferent on all but one subset $\{a, b\}$ of a triple $\{a, b, c\}$. For simplicity, we state this property for additively separable aggregators.

Definition 6. *We say $S \in \mathcal{S}(\{a, b, c, T\})$ is a (δ, ε)-approximate triple-basis for f with respect to $\{a, b, c\}$ if $f(a, \{a, b\}, \{a, b, c\}, S) = f(b, \{a, b\}, \{a, b, c\}, S) + \delta$ while for all other $A \subseteq \{a, b, c\}$ and $x, y \in A$, $|f(x, A, \{a, b, c\}, S) - f(y, A, \{a, b, c\}, S)| < \varepsilon$.*

That is, S is a (δ, ε)-approximate triple basis for f if given choice set $\{a, b\}$ the aggregated utility of U for a is exactly δ higher than the aggregated utility of b, while S is ε-indifferent among all alternatives given every other choice set. We say that an aggregator f is *approximately triple-solvable with k selves* if there is $\overline{\delta} > 0$ such that exists a (δ, ε)-approximate triple-basis with k selves for every $\delta < \overline{\delta}$ and $\varepsilon > 0$. That is, for approximate triple-solvability we do not require that the triple basis is exactly indifferent between all elements in choice sets other than $\{a, b\}$, only that they can be arbitrarily close to being indifferent.

For some aggregators, approximate triple-solvability yields a triple-basis with fewer utility functions. Indeed, consider an aggregator of the form $f(a, A, S) = \sum_{(u,t)\in S} h(\max_{a' \in A} u(a'))u(a)$, where $\lim_{x\to\infty} h(x)x = 0$. Under such an aggregator, the presence of an alternative with a very high utility level under one 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 s: let the self s have $u(a)$ and $u(b)$ such that $f(a, \{a, b\}, \{a, b, c\}, s) - f(b, \{a, b\}, \{a, b, c\}, s) = \delta$ (for small enough δ this is always possible), and let $u(c)$ be high enough so s is ε-indifferent between any two elements given sets containing c. The following is a corollary of the proof of Theorem 2.

Corollary 4. *Suppose $f \in \mathcal{F}$ is approximately triple-solvable with k_f selves. Then, for any finite set of alternatives X, and any choice function $c : P(X) \to X$ that exhibits at most $\frac{n-1}{k_f}$ IIA-violations, f can rationalize c with n utility functions.*

Proof. The only difference compared to the proof of Theorem 2 is the construction of the utility functions. Recall the definition of $(I_j)_{j=1,\dots,j^*}$ from the proof of Theorem 2. Let $\delta_1 \in (0, \overline{\delta})$.

Define iteratively δ_j for $j \in \{2, ..., j^* + 1\}$ such that $\delta_j \in (0, \frac{\delta_{j-1}}{\text{IIA}(c)+1})$. Define a self s^X such that s^X is δ_{j^*+1}-indifferent and the preference ordering of the self is $c(X) \succ c(X \setminus \{c(X)\}) \succ ...$ Let $\delta^{**} = \min_{x \neq y \in X,\ A \ni x,y} |f(x, A, X, u^X)| - |f(y, A, X, u^X)|$. Finally, let $\varepsilon \in (0, \frac{\delta^{**}}{|X|})$. Then for every $j \in \{1, ..., j^*)$ and $A \in I_j$ construct a collection $S^A \in \mathcal{U}(X)$ the following way: take a (δ_j, ε)-approximate triple-basis S, and define S^A by defining, for each s_i a utility function u_i^A as follows. We let $u^i(A)(x)$ equal: $u_i(a)$ if $x = c(A)$; $u_i(b)$ if $x \in A \setminus \{c(A)\}$; $u_i(c)$ if $x \in X \setminus A$. Proving the collection of s^X and S^A for each $A \in \bigcup_{j=1}^{j^*} I_j$ rationalizes c is analogous to the proof in Theorem 2. ∎

D Relaxing P6

Our main results can be extended to aggregators violating P6, that is, to aggregators that depend in a nontrivial way on alternatives unavailable in a given choice set. However, the appropriate definition of triple-solvability is more complicated. The main complication arising in the absence of P6 is that triple-solvability needs to be defined on a general X, as opposed to just a triple $\{a, b, c\}$. It is convenient to introduce the following notation: for any triple $\{a, b, c\}$, any basic set of alternatives $X \supset \{a, b, c\}$, and any self $s = (u, t)$ defined on $\{a, b, c\}$, define the set $E(s, X) = \{(\widehat{u} : X \to \{u(a), u(b), u(c)\}, t) | \widehat{u}(x) = u(x)\ \forall\ x \in \{a, b, c\}\}$. In words, $E(s, X)$ is the set of extensions of the self from $\{a, b, c\}$ to X for which each element in $X/\{a, b, c\}$ has the same utility as a, b or c. For any $S = \langle s_1, ..., s_m \rangle \in \mathcal{S}(\{a, b, c\}, T)$, let $E(S, X) = \{\langle \widehat{s}_1, ..., \widehat{s}_m \rangle | \widehat{s}_i \in E(s_i, X)\ \forall i\}$.

Definition 7. *We say $S \in \mathcal{S}(\{a, b, c\}, T)$ is a universal triple-basis for f if for any $X \supset \{a, b, c\}$ the following holds: for all $\hat{S} \in E(S, X)$, $f(a, \{a, b\}, X, \hat{S}) > f(b, \{a, b\}, X, \hat{S})$, and $f(\cdot, A, X, \hat{S})$ is constant for all other $A \subseteq \{a, b, c\}$.*

A universal triple-basis solves the triple $\{a, b, c\}$ whenever the utilities of unattainable elements don't differ from utilities of elements in $\{a, b, c\}$, for all selves in the triple-basis. An aggregator f is *universally triple-solvable* if the following condition is satisfied.

Universal triple-solvability. There is a triple $\{a, b, c\}$ and $k \in Z_+$ such that for all $\delta > 0$ there is a δ-indifferent $S \in \mathcal{S}^k(\{a, b, c\}, T)$ which is a universal triple-basis for f with respect to $\{a, b, c\}$.

It is easy to see that for aggregators satisfying P6, universal triple-solvability is equivalent to triple-solvability. If f satisfying P1-P5 is universally triple-solvable with k selves, then the same construction can be applied as in the proof of Theorem 2 to obtain an analogous lower bound on the set of choice functions that f can rationalize with a given group size. The proof of this result is analogous to the proof of Theorem 2 and hence omitted.

Corollary 5. *Suppose f satisfies P1-P5 and is universally triple-solvable wrt to X with k_f selves. Then, using n selves, f can rationalize any choice function, on any grand set of alternatives X, that exhibits at most $\frac{n-1}{k_f}$ IIA-violations.*