RATIONALIZING CHOICE WITH MULTI-SELF MODELS

By

Attila Ambrus and Kareen Rozen

July 2008
Updated May 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1670

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Rationalizing Choice with Multi-Self Models*

Attila Ambrus†
Duke University

Kareen Rozen‡
Yale University

May 2012

Abstract

This paper studies a class of multi-self decision-making models proposed in economics, psychology, and marketing. In this class, choices arise from the set-dependent aggregation of a collection of utility functions, where the aggregation procedure satisfies some simple properties. We propose a method for characterizing the extent of irrationality in a choice behavior, and use this measure to provide a lower bound on the set of choice behaviors that can be rationalized with n utility functions. Under an additional assumption (scale-invariance), we show that generically at most five "reasons" are needed for every "mistake."

JEL Codes: D11, D13, D71
Keywords: Multi-self models, index of irrationality, IIA violations, rationalizability

*First version: April 2008. Previous versions of this paper were distributed under the title "Revealed Conflicting Preferences." We are especially grateful to Ben Polak and Larry Samuelson, as well as 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, Ariel Rubinstein, Rani Spiegler, and Tomasz Strzalecki for their helpful suggestions. We thank seminar audiences at Brown, Harvard, MIT, Montreal, NYU, UCL, Yale and the North American Summer Meeting of the Econometric Society for insightful comments.

†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

Suppose you see a group or individual decision-maker exhibiting choice behavior which is inconsistent with standard utility maximization. It is presumed that this choice behavior arises from the aggregation of multiple utility functions (e.g., corresponding to different individuals, selves, or rationales) using some given method of aggregation. How many utility functions would be sufficient to explain this behavior? Conversely, given a method of aggregation, can one identify choice behaviors that are explainable with a certain number of utility functions? This paper studies these questions within a framework of utility aggregation. Consider the following examples of aggregation methods (included in this framework), where U is a collection of utility functions over a space of alternatives X.

Example 1 - Utilitarianism. The aggregate utility of an alternative a in a choice set A is given by $\sum_{u \in U} u(a)$. 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 $\sum_{u \in U} \Phi(\max_{b \in A} u(b) - \min_{b \in A} u(b)) u(a)$, where the contribution of $u(a)$ to the aggregate utility depends via Φ on the range of u over choice set A. For binary choice sets, this 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. For larger choice sets, and when Φ is the identity, this is the *focus-weighted* model of Kőszegi and Szeidl (2012).

Example 3 - Costly self-control. Suppose a decision-maker's long-run self can exert costly self-control over multiple short-run selves, where the cost of such self-control depends on how tempting the alternatives are. The aggregate utility of an alternative a in a choice set A is given by $u^{\mathrm{lr}}(a) - \gamma \sum_{u^{\mathrm{sr}} \in U} [\max_{a' \in A} u^{\mathrm{sr}}(a') - u^{\mathrm{sr}}(a)]^{\psi}$, where u^{lr} and the u^{sr}'s correspond to the long-run self and the short run selves, respectively, and γ, ψ are self-control parameters. If there is only one short-run self, this model is the reduced-form version of Fudenberg and Levine (2006)'s dual-self impulse control model.

Only rational choice can be explained by Example 1, regardless of the number of utility functions being aggregated. Examples 2 and 3, however, belong to a literature on multi-self or multi-utility decision-making which has surged in response to evidence against rational choice behavior. Proposed models include May (1954), Kalai, Rubinstein and Spiegler (2002), Fudenberg and Levine (2006), Manzini and Mariotti (2007), Green and Hojman (2009), and

Kőszegi and Szeidl (2012) in economics, where selves are often seen as rationales or manifestations of temptation and self control processes; Kivetz, Netzer and Srinivasan (2004) and Orhun (2009) in marketing, where selves are different criteria for evaluating products; and Tversky (1969), Shafir, Simonson and Tversky (1993) and Tversky and Simonson (1993) in psychology, where selves are different motivational systems. A first feature of these papers is that they explain "irrational" choice behavior as arising from the aggregation of multiple objectives.[1] This accords with the view in psychology that "the singular self is a hypothetical construct, an umbrella under which experiences are organized along various dimensions" (Lachmann 1996). A second important feature is that, like in Examples 2 and 3, the contribution of a given self to aggregate utility may depend on the choice set; this is related to the idea in psychology that the self "is fluid in that it shifts in different contexts" (Lachmann 1996). As seen from Example 1, both features may be needed to explain irrational choice behavior.

The present paper studies a framework of utility aggregation that encompasses many multi-self models in the literature, including Examples 1-3. An *aggregator*, or model of aggregation, acts on a collection of utility functions (selves, rationales, or members of a group) to assign an aggregate utility to each alternative in each choice set. The decision-maker has a model of aggregation and a collection of utility functions which he uses to select, from each choice set, the alternative which maximizes aggregate utility. The class of aggregation models studied here has several basic features exemplified by Examples 1-3. First, the aggregate utility of an alternative may depend on the choice set only through the set of utility levels potentially attained by the different selves. Second, if two collections of utility functions each assign higher aggregate utility to one alternative than another alternative, then the aggregate utility under the combined collection of utility functions preserves that ranking. Third, we study aggregators possessing some weak continuity properties ensuring that (1) the cardinality or "intensity" of a self's preference may affect choice and (2) aggregation is non-dictatorial.[2]

[1] An expanded shortlist of the multiple-selves or multiple-utility literature includes Benabou and Pycia (2002), Masatlioglu and Ok (2005), Evren and Ok (2007), and Chatterjee and Krishna (forthcoming). This literature is also related to the application of social choice tools in multi-criteria decision problems, as in Arrow and Raynaud (1986), and is related more generally to the theory of multiattribute utility (see Keeney and Raiffa (1993)). Another approach, developed in Bernheim and Rangel (2007) and Salant and Rubinstein (2008), allows for context-dependence by considering extended choice situations where behavior can depend on unspecified *ancillary conditions* or *frames*. While information effects can explain some context dependence (Sen (1993), Kochov (2007), Kamenica (2008)), they cannot explain many systematic violations of IIA (Tversky and Simonson (1993)).

[2] Using cardinal information in the utility functions is a common feature of models of household and

In this paper, choice behavior is captured by a choice function. We say that a choice function is *rationalized* by a model of aggregation and a collection of utility functions if the choice from each set is the unique maximizer of aggregate utility. We show that to answer the question of how many utility functions would be sufficient to rationalize a choice function under a given model of aggregation, one may simply multiply two numbers. The first one is the choice function's *index of irrationality*, as given by our proposed accounting procedure for violations of IIA (independence of irrelevant alternatives). The second one is a number inherent to the method of aggregation itself, called the *proportionality constant of the aggregator*. The proportionality constant, which is independent of both the choice function and alternative space in question, is found by examining the aggregator's behavior over an arbitrary three-element set. For utilitarianism, the proportionality constant is infinite (irrational behavior cannot be explained); for Examples 2 and 3, as well as a class of scale-invariant aggregation models, we show that the proportionality constant is uniformly bounded by five (at most five reasons are needed to justify each mistake); for another class of aggregation models having a "populist" flavor, we show that the proportionality constant is one (at most one reason is needed to justify each mistake). Given any model in our class, our results thus identify a lower bound on the set of choice functions that can be rationalized with a fixed number of utility functions. Our results thus identify models for which it is important to impose *a priori* restrictions on the number of selves (such as positing a "dual self" model) or the number of attributes of a good which are under consideration — else the theory may not have any testable implications. This need not be the case outside the class of models studied here: for example, the models ofManzini and Mariotti (2007) and de Clippel and Eliaz (2012) can explain only certain types of irrational behaviors even when using arbitrarily many rationales.

In the models treated by this paper, it can happen that the choice is not the most preferred alternative according to any of the utility functions, but is the *best compromise*, in the sense that it maximizes aggregate utility. Because of this feature, it may be difficult to tell *a priori* whether it is possible to construct a rationalization of a choice function. A corollary of our results is a universal procedure for constructing a rationalization (of any choice function), using any method of aggregation having a finite constant of proportionality.

intrapersonal decision-making, where the intensities of preferences should be comparable and may play an important role. Without cardinal information, it would be hard, for example, to model a home-buyer who believes neighborhood safety is more important than proximity to work, but would be willing to trade a small enough degree of safety for a shorter commute.

1.1 Related literature

A different concept of rationalization in studied by Kalai, Rubinstein, and Spiegler (2002), who connect the complexity of a rationalization — which they measure by the number of selves — to the size of the space of alternatives (as opposed to the number of IIA violations). They say that a collection of strict preference relations rationalizes a choice function if the choice from each set is optimal for at least one of the preference relations. In this view, each (ordinal) self serves as a dictator for some subset of choices. They find that to explain a choice function defined on an alternative space with n elements, it suffices to posit n different selves. An ordinal and dictatorial model as above would not satisfy the properties of the models studied here. However, one may reinterpret the Kalai et al. (2002) framework as a model of aggregation using this paper's definition of rationalization, where each utility function in the model is assigned some choice sets over which it is the sole contributor to aggregate utility (i.e., dictator). Under that interpretation, to rationalize a particular choice function, one would have to assign to each utility function the sets over which it acts as dictator, which amounts to modifying the method of aggregation itself. By contrast, this paper is interested in studying the set of behaviors rationalizable by a fixed aggregator.[3] Because the aggregator is fixed, and utility functions may have some contribution to aggregate utility for every choice set (through a potentially complex functional form), constructing a rationalization for models within our framework is a more complicated matter than in Kalai et al. (2002). In contrast to their main result, for any given model of aggregation, our universal method for constructing a rationalization uses the same number of selves for any choice function which has n violations of IIA according to our index – independently of the size of the space of alternatives.

Our results complement those in the household choice literature, such as Browning and Chiappori (1998) and Chiappori and Ekeland (2006). That literature differs in several ways from the present setting. In view of evidence that household demand cannot arise from the maximization of a single utility, they examine microeconomic implications of Pareto-efficient choice, where each member of the household is a utility maximizer. Given a household demand function over m goods (which may be viewed as an incomplete choice function), they ask when there exist n utility functions $\{u_i\}_{i=1}^n$ and a continuously differentiable function μ of prices and income such that the demand arises from the weighted utilitarian maximization of $\sum_{u\in U}\mu(\text{price,income})u(\cdot)$ given the budget set (i.e., weights and preferences vary independently). Browning and Chiappori (1998) show that if there are n goods, then any demand

[3] Under another interpretation, they would never have to modify the model of aggregation if it always required as many selves as there are choice sets. But then the number of selves is fixed at $2^n - n - 1$.

data can be explained by an $(n-1)$-person household. Chiappori and Ekeland (2006) show that to explain a given demand function using n people, it is necessary and sufficient that the rank of a certain matrix in a pseudo-Slutsky matrix decomposition be $n-1$, though without further restrictions there can be a continuum of explanatory n-person models. While the above papers assume that the modeler does not know the household's weighting rule, the current paper addresses the question of rationalization with a fixed aggregator. Moreover, the weight a utility function receives in the present framework depends on utility levels but not directly on price or income; hence weights and preferences cannot vary independently in our framework.

There are several recent contributions to the literature on multi-self decision-making which mostly focus on a different set of questions. Of these, most related is Green and Hojman (2009), who also study a class of aggregation methods. Because they model a DM as a probability distribution over all possible ordinal preference rankings, their framework is difficult to compare to models of multi-self decision-making with a discrete number of cardinal selves, but is related to models in the voting literature (e.g., Saari 1999). Extending results from that literature, they show that if choice is determined by a voting rule satisfying a monotonicity property, then their model can explain any choice behavior.[4] The rest of their paper focuses on welfare analysis. Bernheim and Rangel (2007) and Chambers and Hayashi (2008) also focus on welfare analysis given choices contradicting rational decision-making. Other related work includes Manzini and Mariotti (2007), Masatlioglu and Nakajima (2007) and Cherepanov, Feddersen and Sandroni (2010), who consider sequential application of multiple rationales to eliminate alternatives, a process they show can rationalize certain choice functions.

1.2 Organization of the paper

In Section 2 we formalize the basic framework of utility aggregation, the class of models studied here, and our concept of rationalization. In Section 3 we describe our index of irrationality. We begin by introducing our results in Section 4 for the special class of scale invariant models of aggregation. Our rationalizability result, as well as a sketch of proof describing our general procedure for constructing rationalizations, are given in Section 5. We present several extensions in Section 6. An extension of our notation for the basic

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

framework and results to include type-dependent aggregation is presented in Supplementary Appendix A.

2 Framework

An individual's (or group's) choice behavior is observed with respect to a finite space of alternatives X. Let $P(X)$ be the set of nonempty subsets of X. The individual's *choice function* $c : P(X) \to X$ identifies the alternative $c(A) \in A$ chosen from each $A \in P(X)$. A *rationalization* of a choice function consists of two components, a collection of *utility functions* U and an *aggregator* f that combines these utilities into an aggregate utility function, in a way that possibly depends on the choice set. Viewed as a multi-self model, these utility functions represent a decision-maker's conflicting motivations or priorities. The aggregator corresponds to the method of sorting out different priorities to come to a decision. To simplify notation, in the main text we restrict attention to a simplified framework in which the aggregator treats all utility functions symmetrically. However, in Supplementary Appendix A we allow nonanonymous aggregation and extend the main results to asymmetric aggregators which treat utility functions differently based on a "type." That feature arises in some models of multi-self decision-making, such as with long and short-run types of selves (for instance, consider Example 3). The extension to types requires more cumbersome notation but no conceptual innovation.

Formally, given the space of alternatives X, a utility function $u : X \to \mathbb{R}$ describes the utility level allocated to each alternative $x \in X$.[5] A *collection of utility functions* U is an unordered list of utility functions. By definition of an unordered list, a collection U can have multiple identical utility functions, and there is no order hierarchy over members of the list. Formally, for a given grand set of alternatives X, a collection U is an element of $\mathcal{U}(X) = \cup_{n=1}^{\infty} \mathcal{U}^n(X)$, where $\mathcal{U}^n(X)$ is the set of all unordered lists of length n of utility functions over X. The number of utility functions in U is denoted by $|U|$, or simply n when no confusion would arise.

[5] Though aggregation in this 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 collection U, then so does any increasing transformation of f. Similarly, if f rationalizes c using the collection U, then $f \circ h^{-1}$ rationalizes c using the collection $h \circ U$, where $h : \mathbb{R} \to \mathbb{R}$ is invertible on the appropriate domain. That is, given any representation U and f, one can obtain an equivalent representation by applying a monotone transformation of utilities in U, if a corresponding transformation is applied to the aggregation function f as well.

An aggregator f specifies an aggregate utility for every alternative a in every choice set A, given a (finite) grand set of alternatives X and any collection U defined over these alternatives.[6] Formally, the domain over which the aggregator f is defined is

$$\{a, A, X, U | X \in \mathcal{X}, U \in \mathcal{U}(X), A \in P(X), a \in A\},$$

where $\mathcal{X}$ is the set of conceivable finite spaces of alternatives. To simplify notation, we will write $f(a, A, U)$ whenever doing so would not cause confusion. Since the choice set A is one of the arguments of the function, f aggregates the utilities in the collection U in a possibly context-dependent way. An aggregation rule may be seen as a particular theory of how members of the collection are activated by choice sets: the aggregator determines the weight each utility function receives on the choice set as a function of its utility levels over the alternatives. We say that a model of aggregation f *rationalizes* a choice function if there is a collection of utility functions such that for every choice set, the alternative that maximizes aggregate utility is the one selected by the choice function.[7]

Definition 1. *A choice function $c : P(X) \to X$ is rationalized by an aggregator f if there is a collection U such that for every choice set $A \in P(X)$, $c(A)$ is the unique maximizer of $f(a, A, U)$ within A.*

In this paper, we study a class of aggregators that contains many multi-self models proposed in the literature – those which are monotonic transformations of an additively separable form, in which the weight each utility function receives depends on the set of utility levels it attains on the choice set. Formally, the class of models $\mathcal{F}$ contains all models of the form

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

where h is an increasing transformation, and where g – which evaluates each alternative $a \in A$ based on the set of utility values u takes on A – satisfies the following two properties. First, g satisfies *consistency*: $g(a, \{u(a')\}_{a' \in A}) \geq g(b, \{u(a')\}_{a' \in A})$ whenever $u(a) \geq u(b)$. This is a minimal consistency requirement, ensuring that g preserves the ranking of u. Second, g satisfies *neutrality*: $g(a, \{u(a')\}_{a' \in A}) = g(\pi \circ a, \{u(\pi \circ a')\}_{a' \in A})$ for any permutation $\pi : X \to$

[6] We could permit aggregators with restricted domains: for convex $\hat{\mathbb{R}}^X \subset \mathbb{R}^X$, let $\mathcal{U}^n = \times_{i=1}^n \hat{\mathbb{R}}^X$.

[7] The underlying model 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.

X. This says the treatment of alternatives depends on utilities, not names. As described in the Appendix, models in the class $\mathcal{F}$ satisfy six axiomatic properties on which our results are based.

The class of models $\mathcal{F}$ contains many familiar examples. The model of utilitarianism in Example 1 is one such instance, as is the generalization of the additive difference model of Tversky (1969) in Example 2.[8] 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, as in the focus-weighted model of Kőszegi and Szeidl (2012) (where each utility function corresponds to a dimension of a good, and Φ is the identity). 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 of alternatives. Example 3, the dual-self model of Fudenberg and Levine (2006), does not exactly fit into the class $\mathcal{F}$ because the utility function corresponding to the long-run self in the functional form $u^{\mathrm{lr}}(a) - \gamma \sum_{u^{\mathrm{sr}} \in U} [\max\limits_{a' \in A} u^{\mathrm{sr}}(a') - u^{\mathrm{sr}}(a)]^{\psi}$ is treated differently than the short-run selves; however, the summation corresponding to the short-run selves is in $\mathcal{F}$, as is the part corresponding to the long-run self. Our results extend in this case and others with nonanonymous aggregation; see Supplementary Appendix A. Finally, we mention below some other models in $\mathcal{F}$ which are equivalent or closely related to ones in the existing literature.[9]

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 $m\Bigg(\sum_{u \in U} u(a) \Bigg) + \sum_{u \in U} \ell \Bigg(u(a) - r(\{u(a')\}_{a' \in A}) \Bigg)$, where $r(\cdot)$ determines the reference point against which $u(a)$ is evaluated for each utility function u; $m(\cdot)$ 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. Kivetz et al. (2004) consider goods

[8] Tversky (1969) accounts for potentially 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 $u(a) - u(b)$.

[9] Two additional aggregators are studied in Supplementary Appendix 7. where we show how to rationalize two simple choice procedures discussed in Kalai et al. (2002): the median procedure and the second-best procedure. In particular, Kalai et al. (2002) show that within their framework, the number of selves needed to rationalize these choice procedures becomes unbounded as the alternative space grows large. We show they can be rationalized in our framework with two selves, regardless of the size of the alternative space.

(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. Similarly 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.

Example 5 - Nash bargaining solution with an endogenous disagreement point. The aggregate utility of an alternative a in a choice set A is $\prod_{u \in U}(\kappa + u(a) - \min_{a' \in A} u(a'))$, where κ is any positive constant to ensure each term is strictly positive.

Example 5, 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).

These and other models in $\mathcal{F}$ share some prominent characteristics. First, as in many existing models of household and multi-self decision-making (and in expected utility), aggregation depends on cardinal information in the utility functions. Building on cardinality, these models offer the possibility of compromise. This is a defining feature of the models of household choice, which are interested in efficient outcomes arising from an unmodeled bargaining process. As opposed to Kalai et al. (2002), but in accordance with others (e.g., Tversky (1969), Tversky and Kahneman (1991), Kivetz et al. (2004), Fudenberg and Levine (2006), Green and Hojman (2009)), the utility functions in these models have some contribution to aggregate utility on every choice set.[10] Finally, the weight allocated to a utility function may depend on how it evaluates the options in the choice set. For example, in Fudenberg and Levine (2006), a long-run self must exert more costly self control when more appealing options are available. Such models can also capture the behavior in Shafir et al. (1993), where the primary rationales for purchasing depend on the attributes of available products. As seen from Example 1 (utilitarianism), the weights allocated to utility functions must depend on the choice set in some way in order to generate irrational behavior.

[10] Psychologists believe that a fluid form of compromise among selves is necessary for healthy behavior. This is as opposed to disassociated selves (i.e., overly autonomous selves), or a high self-concept differentiation (a lack of interrelatedness of selves across contexts) both of which are connected to pathological or unhealthy behavior; see Power (2007), Donahue, Robins, Roberts and John (1993), and Mitchell (1993).

3 An index of irrationality

What kinds of behavior can an aggregator rationalize? Consider one of the simplest types of aggregators, the model of utilitarianism: $f(a, A, U) = \sum_{u \in U} u(a)$. The only choice function that utilitarianism can rationalize is rational choice, that is, choice which satisfies the Independence of Irrelevant Alternatives (IIA). 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. A non-utilitarian model of aggregation, however, might be able to rationalize a choice function that violates IIA. To identify choice functions that deviate from rationality but are rationalizable, this paper provides an index of irrationality based on 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 determine whether a set (or more precisely, the choice from that set) causes an IIA violation. Consequently, there are different ways to define the number of violations. One possible measure would be based on the following characterization of an IIA violation.

Characterization 1. *A set A causes an IIA violation given $c(\cdot)$ if there exists a set $B \supset A$ such that $c(B) \in A \setminus \{c(A)\}$.*

To illustrate, consider $X = \{w, x, y, z\}$ and the choice function $c(\cdot)$ given by:

$$c(\{w, x, y, z\}) = w,$$
$$c(\{w, x, y\}) = c(\{w, y, z\}) = y,\ c(\{w, x, z\}) = w,\ c(\{x, y, z\}) = x,$$
$$c(\{w, x\}) = c(\{w, z\}) = w,\ c(\{w, y\}) = c(\{y, z\}) = y,\ c(\{x, y\}) = c(\{x, z\}) = x.$$

The sets $\{w, x, y\}$, $\{w, y, z\}$, and $\{w, y\}$ are IIA violations under this characterization, because IIA dictates that if $c(\{w, x, y, z\}) = w$ then the choice from those sets should be w. However, notice that choosing y from $\{w, y\}$ is consistent with IIA in view of the fact that y is chosen from $\{w, x, y\}$ and $\{w, y, z\}$. Consequently, one might want to view A as an IIA violation *only if* its choice contradicts that of the *first* superset having an IIA implication for A. This idea is formalized below.

Characterization 2. *A set A causes an IIA violation given $c(\cdot)$ if there exists a set $B \supset A$*

such that $c(B) \in A \setminus \{c(A)\}$ and for every A' such that $A \subset A' \subset B$, we have $c(A') \notin A$.

Under Characterization 2, only $\{w, x, y\}$ and $\{w, y, z\}$ are violations in the example above; in view of those choices, IIA dictates picking y from $\{w, y\}$. Thus, the second characterization goes further than the first in viewing each choice set A causing a violation as a *regime change*; that is, as having IIA implications in all subsets of A in which $c(A)$ is contained. As evidenced by the following example, one could take the characterization of IIA violations as regime changes even further:

$$c(\{w, x, y, z\}) = w,$$
$$c(\{w, x, y\}) = c(\{w, x, z\}) = w,\ c(\{w, y, z\}) = y,\ c(\{x, y, z\}) = x,$$
$$c(\{w, x\}) = c(\{w, z\}) = w,\ c(\{w, y\}) = c(\{y, z\}) = y,\ c(\{x, y\}) = c(\{x, z\}) = x.$$

IIA implies that if w is chosen from $\{w, x, y, z\}$, then it should also be chosen from $\{w, y, z\}$ and $\{w, y\}$. At the same time, the choice of y from $\{w, y\}$ is implied by the choice of y from $\{w, y, z\}$ (which itself contradicts the choice of $\{w, x, y, z\}$), but not by the choice of w from $\{w, x, y\}$. Under Characterization 2, $\{w, y\}$ is considered a violation – even though y is implied by the IIA violation $\{w, y, z\}$, there is no choice set in between $\{w, y\}$ and $\{w, x, y\}$. Consider now the view that each IIA violation is a regime change which has implications for all subsets in which the choice is contained. Then, the choice of y from $\{w, y, z\}$ implies that y should be chosen from any subset of $\{w, y, z\}$ in which it is contained – in particular, y should be chosen from $\{w, y\}$. Moreover, no other regime change has a contradictory implication for $\{w, y\}$. Consequently, under the following characterization, which refines the previous two, the only set causing a violation is $\{w, y, z\}$. The characterization is iterative, starting from X, then examining sets of size $|X| - 1$, etc., until reaching sets of size two.

Characterization 3. *The set X does not cause an IIA violation. Inductively, for a set $A \subset X$, let $\mathcal{V}(A)$ be the set of smallest supersets of A which cause an IIA violation and whose choice is contained in A. Then we say that A causes an IIA violation if*

(1) There exists B such that $A \subset B$, $c(B) \in A \setminus \{c(A)\}$ and for every A' such that $A \subset A' \subset B$, $c(A') \notin A$; and

(2) If $\mathcal{V}(A) \neq \emptyset$ then there exists $B' \in \mathcal{V}(A)$ such that $c(B') \neq c(A)$.

Condition (1) is simply Characterization 2. The refinement in condition (2) ensures a set is not considered a violation if its choice is implied by the previous regime changes. For any choice function, the number of sets causing an IIA violation is smallest under Characterization 3 and highest under Characterization 1. We define the index of irrationality as follows.

Definition 2 (Index of Irrationality)**.** *The index of irrationality of a choice function* $c(\cdot)$ *is given by* $\mathrm{II}(c) = \#\{A \in P(X) \mid A$ *causes an IIA violation under Characterization 3*$\}$.

Because the results in this paper determine how many utility functions would be sufficient to explain a choice behavior with a given index of irrationality, the result is tighter the smaller is the index of irrationality used. Another possible measure would be the minimal number of sets where the choice function must be changed to make it rational; this measure, however, is not comparable with our own – it can be either larger or smaller than the index above.[11]

4 Scale-invariant models

We begin by introducing our results for a special class of models in $\mathcal{F}$ which satisfy a scale-invariance property. Models in $\mathcal{F}$ are ordinally equivalent to the form $\sum_{u \in U} g(a, \{u(a')\}_{a' \in A}))$. We say that $f \in \mathcal{F}$ is in the class of models $\mathcal{F}^*$ if $g(a, \{\alpha u(a')\}_{a' \in A}) = \phi(\alpha) g(a, \{u(a')\}_{a' \in A})$ for any $\alpha \in \mathbb{R}$ and some invertible and odd $\phi : \mathbb{R} \to \mathbb{R}$. This says the unit in which preference intensity 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 $\sum_{u \in U} (u(a) - \text{mean } u(A))^\rho$, 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

[11] Indeed, suppose that pairwise choices exhibit the transitive ranking a preferred to b preferred to c. Under this paper's 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 in the example above. The alternative measure counts two violations, while this paper's measure counts one.

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 sup metric through which genericity is defined, note that by scale invariance, there is a natural bijection between (1) aggregators 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) Fix any aggregator $f \in \mathcal{F}^*$ and three alternatives $x, y, z \in X$. If f 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 f can rationalize any choice function c defined over X.

(ii) The set of aggregators in $\mathcal{F}^*$ that can rationalize any choice function c using at most $1 + 5 \cdot II(c)$ 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 even a small amount of irrationality can be explained (only two types of irrational behavior out of the three possible types of irrational behavior over three alternatives), an additive and scale-invariant model can rationalize any choice behavior with sufficiently many utility functions. 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 exhibits an index of irrationality of at most $\frac{n-1}{5}$. 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 individuals or selves in order to generate a refutable theory.

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 utility function defined over a triple $\{x, y, z\}$ for which f "stretches" utility differences over pairs, $f(x, \{x, z\}, u) - f(z, \{x, z\}) \neq f(x, \{x, y\}, u) - f(y, \{x, y\}, u) + f(y, \{y, z\}, u) - f(y, \{y, z\}, u)$; and for which f's evaluation of alternatives in the triple is not fixed by the pairwise rankings, $f(x, \{x, y, z\}, u) - f(y, \{x, y, z\}, u) + f(x, \{x, y, z\}, u) - f(z, \{x, y, z\}, u) \neq f(x, \{x, y\}, u) - f(y, \{x, y\}, u) + f(x, \{x, z\}, u) - f(z, \{x, z\}, u)$. For example, the utility function $u(y) = 4 > u(z) = 2 > u(x) = 1$ shows that the model in Example 4 using an arithmetic mean is in the generic class. By contrast, utilitarianism and generalizations of the form $f(a, A, u) = u(a) + h(A)$, 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

In this section we present our main result, which determines how many utility functions would be sufficient to explain a choice function based on its index of irrationality. Before doing so, we begin with an illustrative example.

5.1 Illustration

Recall the model in Example 2, where the aggregate utility of an alternative $a \in A$ is

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

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, consider the following collection of utility functions $U = (u_1, u_2, u_3, u_4, u_5)$ 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 the aggregator selects a from the choice set $\{a, b\}$. Observe that $f(a, \{a, b\}, U) = 4\Phi(2) + \Phi(1)$ and $f(b, \{a, b\}, U) = 2\Phi(2) + 3\Phi(1)$. Hence $f(a, \{a, b\}, U) > f(b, \{a, b\}, U)$ since $\Phi(2) > \Phi(1)$. For any other choice set, the aggregator assigns equal utility to all alternatives:

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

That is, under the collection U, 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 U defined on a three-alternative set $\{a, b, c\}$ a *triple-basis* for this aggregator f. As we now show, triple-bases can serve as basic building blocks for rationalizations of choice functions on arbitrary spaces of alternatives. Recall the example choice function $c(\cdot)$, defined on $X = \{w, x, y, z\}$, from Section 3:

$$\begin{gathered}
c(\{w, x, y, z\}) = w, \\
c(\{w, x, y\}) = c(\{w, x, z\}) = w, \; c(\{w, y, z\}) = y, \; c(\{x, y, z\}) = x, \\
c(\{w, x\}) = c(\{w, z\}) = w, \; c(\{w, y\}) = c(\{y, z\}) = y, \; c(\{x, y\}) = c(\{x, z\}) = x.
\end{gathered}$$

This choice function's index of irrationality is equal to one. Using the triple basis above, we construct the following collection $U^{\{w,y,z\}} = (u_1, u_2, u_3, u_4, u_5)$, defined on X:

u_1		u_2		u_3		u_4		u_5	
w, z	2	w, z	2	x	2	x, y	2	y	2
x	1	y	1	w, z	1	w, z	0	w, x, z	0
y	0	x	0	y	0				

The above utility functions are constructed by letting the choice from $\{w, y, z\}$, which is y, play the role of a in the triple-basis; letting the unchosen alternatives in $\{w, y, z\}$, which are w and z, play the role of b; and finally, letting the alternatives not in $\{w, y, z\}$, which is only x here, play the role of c.

Using $U^{\{w,y,z\}}$, how does f evaluate the alternatives in each choice set $A \subseteq X$? Since x has the same utility as c in the calculations above, it is easy to set that $f(\cdot, A, U^{\{w,y,z\}})$ is constant for any set A containing x. Since the unchosen alternatives w, z in $\{w, y, z\}$ are given equal utilities, $f(\cdot, A, U^{\{w,y,z\}})$ is constant on $\{w, z\}$. Finally, since y plays the role of a, and w, z play the role of b, the previous calculations imply that for any $A \subseteq \{w, y, z\}$ containing y, we have

$$f(y, A, U^{\{w,y,z\}}) > f(\tilde{y}, A, U^{\{w,y,z\}}) \text{ for all } \tilde{y} \in A \setminus \{y\}.$$

Therefore, the utility functions in $U^{\{w,y,z\}}$ rationalize the choice from any set $A \subset \{w, y, z\}$ which contains y, and have no impact on any other choice set.

Since the collection $U^{\{w,y,z\}}$ has implications only for the IIA violation $\{w, y, z\}$ and its subsets, one needs an additional utility function to rationalize the remaining "rational" choices. We construct a final utility function u^* whose utility range is sufficiently small to not overturn any strict preferences induced from $U^{\{w,y,z\}}$, and which has the ranking $u^*(w) > u^*(x) > u^*(y) > u^*(z)$ derived from standard revealed preference; that is, from observing $w = c(X)$, $x = c(X \setminus \{w\})$, and $y = c(X \setminus \{w, x\})$. By construction, the utility functions $(u^*, U^{\{w,y,z\}})$ rationalize $c(\cdot)$.

5.2 Rationalizability result

Observe that the triple basis U 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 utility functions in U can rationalize 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 3. *Given a triple $\{a, b, c\}$ and model of aggregation f, the collection $U \in \mathcal{U}(\{a, b, c\})$ is a triple-basis if $f(a, \{a, b\}, U) > f(b, \{a, b\}, U)$ and $f(\cdot, A, U)$ is constant for all other $A \subseteq \{a, b, c\}$. The aggregator f is triple-solvable with k utility functions if for every $\delta > 0$, there is a triple-basis $U \in \mathcal{U}^k(\{a, b, c\})$ with $\max_{a,b \in A, A \subseteq \{a,b,c\}, u \in U} |f(a, A, u) - f(b, A, u)| < \delta$*

Given an aggregator, it is easy to check for the existence of a triple-basis. Indeed, triple bases can be found for the aggregators featured earlier.[12] For scale-invariant aggregators, which satisfy the property that measuring utilities in a different unit does not change the ordering implied by the aggregator, checking the property is particularly simple, since it then suffices to construct one triple-basis which can be scaled as needed. 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 aggregators featured here, and in fact the class of aggregators $\mathcal{F}^*$ generically satisfies 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 $f \in \mathcal{F}$ is triple-solvable with k_f utility functions. Then, for any choice function c, defined on any finite grand set of alternatives X, no more than $1 + k_f \cdot II(c)$ utility functions are needed to rationalize c.*

We sketch below the proof of Theorem 2, describing our general rationalization method. Note that an alternative statement of the result is as follows: using n utility functions, f can

[12] 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): $u_1(a) = 2.847, u_1(b) = 1, u_1(c) = 7.634$, $u_2(a) = 0, u_2(b) = 4.288, u_2(c) = 1$, $u_3(a) = 6, u_3(b) = -.129, u_3(c) = 4$, $u_4(a) = -4.651, u_4(b) = -.949, u_4(c) = 0$, $u_5(a) = 5, u_5(b) = -1.619, u_5(c) = -15.8$.

rationalize any choice function c, defined on any finite grand set of alternatives X, whose index of irrationality is at most $\frac{n-1}{k_f}$. Hence, the result also gives a lower bound on the set of rationalizable behaviors for a fixed number of utility functions, providing a linear connection between the index of irrationality of a choice function and the degree of freedom in the model (as measured by the number of utility functions).

Note that for each aggregator f, 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 group members in a triple basis). This means that the number of utility functions 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 (under Characterization 3). For each choice set A that does cause an IIA violation, the construction creates a group U^A defined on X, whose utility values match those of a triple basis: $c(A)$ plays the role of the preferred element in the pair, $A \setminus \{c(A)\}$ plays the role of the unchosen element in the pair, and $X \setminus A$ plays the role of the third element. This implies that:

1. $c(A)$ is selected under $f \circ U^A$ from every subset of A in which it is contained; and
2. The group members in U^A cancel each other out under f on every other choice set (that is, on sets not containing $c(A)$ or sets containing some element of $X \setminus A$)

The triple-basis used to generate U^A is "indifferent enough" over the alternatives so that the trickle-down effect of U^A does not overturn the strict preference of previously constructed utility functions (corresponding to other IIA violations). Finally, the construction creates an extra utility function u^*, that is indifferent enough never to overturn any strict preferences from existing utility functions. Using standard revealed preference, u^* allocates the highest utility to $c(X)$, the next highest utility to $X \setminus \{c(X)\}$, and so on.

This procedure constructs $1 + k_f \cdot \mathrm{II}(c)$ utility functions which together rationalize $c(\cdot)$ under the model f. The construction ensures that $c(A)$ is selected from any set causing an IIA violation; one need only check that constructed utility functions do not interfere with choices associated with sets that do not cause IIA violations. To loosely illustrate the idea, consider any nested sequence of choice sets that decreases by one alternative. Given X, or any set which does not cause an IIA violation, all utility functions except u^* are indifferent, hence the preferences of u^* prevail. For the first set of the sequence that contradicts the choice from X, a triple-basis was created with utility functions which overrule u^* and guarantee that the choice from this set under $c(\cdot)$ is the unique f-maximizer (while the other triple-bases created will be indifferent). Similarly, whenever along the sequence there is a set that contradicts the choice of the previous set, another triple-basis was created that overrules all utility functions created in association with larger sets. ∎

It is easy to see that the proposed rationalization procedure can also be modified to generate rationalizations of choice correspondences.

Theorem 1, for scale-invariant aggregators, is proved in five steps. The first is knowing that if f is triple-solvable with k utility functions, we can rationalize any choice function c with $1+k\cdot II(c)$ utility functions. 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 utility functions provides enough degrees of independence to ensure that a triple basis can be constructed.

6 Extensions

6.1 Weakening solvability: one utility function per violation

While triple solvability is a property that is broadly satisfied, it can be seen from our construction that our theorem would still hold under a weaker condition. It suffices that there exist a collection which is arbitrarily close to being indifferent on all but one subset $\{a, b\}$ of a triple $\{a, b, c\}$. We formalize this idea in Supplementary Appendix A, where we extend the notion of a triple-basis to an *approximate triple-basis*. For some aggregators, approximate triple-solvability can yield a triple-basis with drastically fewer utility functions. Indeed, consider an aggregator of the form

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

where $\lim_{x \to \infty} h(x)x = 0$. 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-member* approximate triple-basis u: let $u(a)$ and $u(b)$ such that $f(a, \{a, b\}, \{a, b, c\}, u) - f(b, \{a, b\}, \{a, b, c\}, u) = \delta$ (for small enough δ this is always possible), and let $u(c)$ be high enough so that u is ε-indifferent between any two elements given sets containing c. Theorem 5 in Supplementary Appendix A then implies that only one utility function is needed to rationalize each violation (or alternatively, using n utility function, the aggregator can rationalize all choice functions with index of irrationality up to $n - 1$. This means that for any choice function with one IIA violation, only two selves are needed — which is clearly a tight bound. Note that there are exponentially many choice functions with index of irrationality equal to one: starting from any rational choice function c, choose an arbitrary $A \subset X$, modify the choice to any $a^* \in A \setminus \{c(A)\}$, and make a^* the choice in all subsets of A containing a^*.

6.2 Type-dependent aggregators

The examples of aggregators above all treat utility functions in the same way. However, many models in the existing literature on multi-self decision-making propose methods of aggregation that treat some selves differently than others. For example, Fudenberg and Levine (2006) propose a dual-self impulse control model with a long-run self exerting costly self-control over a short-run self. One way to generalize this aggregator to any number of selves would be to introduce multiple types of short-term temptations, represented by selves $u_1^{\text{sr}}, u_2^{\text{sr}}..., u_n^{\text{sr}}$, as well as one long-run self u^{lr}. Accommodating such type-dependent models of aggregation in our framework requires an extension of the framework and some extra notation, but no conceptual innovation. In particular, the definition of an aggregator must be extended to include a set of possible types, and the definition of a self must be extended to include a type. For ease of exposition, we restricted ourselves to the simplified framework in the main text and present the extension of the framework in Supplementary Appendix A. Our results carry over to the extended framework.

6.3 Systematic IIA violations

Our construction allocates an (approximate) triple-basis for every increment of the index of irrationality. However, there can be IIA violations that are "in the same direction" (that do not contradict each other). In this case, parts of the associated triple-bases in our construction can be combined (or *collapsed*) together to yield tighter bounds. For example, recall the triple-basis for the intensity-weighted aggregator, and fix some alternative a. Every time the choice of a from some set causes an IIA violation, the triple-basis constructed has a utility function u_5 in which a is preferred to $X \setminus \{a\}$, all elements of which are indifferent to each other. Under the intensity-weighted aggregator, all of the u_5's constructed when the choice was a can be collapsed into a utility function.

This effect is particularly pronounced when the triple-basis has only one utility function, as in the approximately triple-solvable aggregators introduced above. Consider the following example: let $x^* \in X$, and let $\succ_1$ and $\succ_2$ be strict orderings on X such that $x \succ_1 x^*$ and $x \succ_2 x^*$ for every $x \in X \setminus \{x^*\}$, and $y \succ_1 x$ for $x, y \in X \setminus \{x^*\}$ if and only if $x \succ_2 y$. Consider a decision-maker who selects the best element according to $\succ_1$ from choice sets not containing x^*, but selects the best element according to $\succ_2$ from choice sets containing x^*. This behavior describes, for example, a customer in a restaurant who chooses the tastiest item from a menu if it does not contain onion rings, while choosing the healthiest item in the presence of onion rings, because they are so greasy as to make the customer feel guilty about his eating habits. The above behavior an index of irrationality of $|X| - 2$.[13] However, these IIA violations do not contradict each other: if the choice from set B contradicts the choice from $A \supset B$, then there is no $B' \subset B$ such that the choice from B' contradicts the

[13] In particular, the sets causing IIA violations are $X \setminus \{x^*\}$, $X \setminus (\{x^*\} \cup \{c(X \setminus \{x^*\})\})$, etc. etc.

choice from B. This can be used to merge all collections of selves into a single collection, reducing the number of selves required to rationalize the customer's choice function. Recall the aggregator introduced in the previous subsection, which was shown to be approximately triple-solvable with a single utility function. Our construction calls for (i) creating one utility function which has the ranking in $\succ_2$; and (ii) creating a utility function for all sets associated with an IIA violation, such that the utility of x^* is sufficiently high that the utility function becomes close enough to indifferent in the presence of x^*, and among the other alternatives allocates the highest utility to the choice from the given set. The latter utility functions can all be collapsed into a single one which agrees with the ranking of $\succ_1$ over $X \setminus \{x^*\}$ (while keeping the utility of x^* at a level that makes it nearly indifferent in the presence of x^*). Our construction implies the above choice function can be rationalized with two selves — which is a tight bound.

7 Discussion

This paper proposes an index of irrationality and examines how many utility functions are sufficient to explain choice behavior, under a large class of existing models of multi-self decision-making. It should not be difficult to incorporate choice correspondences into our results on rationalizability 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). By examining more specific models instead of the broad class of aggregation rules investigated in this paper, sharper predictions on implied choice with a fixed number of utility functions may be possible. As in Kalai et al. (2002), who study a dictatorial model, we provide an upper bound on the number of utility functions required to rationalize a choice behavior. In contrast to their setting, constructing a rationalization using models in the class studied here may be difficult without the aid of the general procedure we provide. Apesteguia and Ballester (2010) have shown that the problem of finding the minimal number of selves needed in the Kalai et al. (2002) model is NP-complete, and we expect finding the minimal number that works in the class of models studied here to be comparably hard. We leave this open problem, as well as extending our framework to dynamic settings, to future research.

Appendix

We prove our results for aggregators satisfying the following properties (it is straightforward to show that these properties are all satisfied by any $f \in \mathcal{F}$). To introduce them, we define one piece of notation: for any two collections of utility functions $U, U' \in \mathcal{U}(X)$, we denote by $\langle U, U' \rangle$ the combined collection $(u_1, \dots, u_{|U|}, u'_1, \dots, u'_{|U'|}) \in \mathcal{U}(X)$.

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

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

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

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

P5 (Profile equivalence)**.** $U(a) = U(\hat{a})$ *implies* $f(b, A \cup \{a\}, U) = f(b, A \cup \{\hat{a}\}, U)$ $\forall\, b \in A$*.*

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

Proof of Theorem 2

For an arbitrary choice function c we will construct a collection of $1 + k \cdot \mathrm{II}(c)$ members which will be shown to rationalize c. This implies the claim in the theorem. In particular, we will construct k members for each set with which an IIA violation is associated, and an extra member for X. Let $I_1 = \{A^1_1, ..., A^1_{i_1}\}$ 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 $I_j = \{A^1_1, ..., A^1_{i_{j+1}}\}$ 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 $\bigcup\limits_{l=1}^{j-1} I_l$ 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 group of k members for each set associated with an IIA violation, starting with sets in I_1. Consider any group of k members $\bar{U}^1 = (\bar{u}^1_1, \dots \bar{u}^1_k)$

that solves the triple $\{a, b, c\}$ (the existence of such a triple follows from triple-solvability). For every $A \subset I_1$, construct now the following group $U^A = (u_1^A, \dots u_k^A)$:

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

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

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

Then by P3 and P6, we know

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

for any $\widetilde{U}_1, ..., \widetilde{U}_m$ group of (exactly) indifferent members. 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 members, there is a δ-indifferent group of k members $\bar{U}^{j+1} = (\bar{u}_1^{j+1}, \dots \bar{u}_k^{j+1})$ that solves the triple $\{a, b, c\}$. For every $A \in I_{j+1}$, construct now the following group $U^A = (u_1^A, \dots u_k^A)$:

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

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

The above procedure generates a group $k \cdot \text{II}(c)$ members in j^* steps. Then by P3 and P4 there is $\delta_{j^*} > 0$ such that for any δ_{j^*}-indifferent u, $f(a, A, X, U_{j^*}) > f(b, A, X, U_{j^*})$ implies $f(a, A, X, (U_{j^*}, u)) > f(b, A, X, (U_{j^*}, u))$. Finally, construct one more member the following way: let $a_1 = c(X)$ and $a_k = c(X \setminus \{a_1, a_2, \dots a_{k-1}\})$ for $2 \leq k \leq n$. Construct $u^* : X \to \mathbb{R}$ such that $u^*(a_1) > u^*(a_2) > \cdots > u^*(a_n)$ and u^* is δ_{j^*}-indifferent. We show $U_c \equiv (U_{j^*}, u^*)$ rationalizes c under aggregator f.

Observation 1. *For any set A with which there is an IIA violation associated, by the construction of U^A and by P1 and P5, $f(a, B, X, U^A) = f(b, B, X, U^A)$ $\forall$ $a, b \in B$ and B such that either $B \setminus A \neq \emptyset$ or $c(A) \notin B$, and $f(c(A), B, X, U^A) > f(b, B, X, U^A) = f(b', B, X, U^A)$ $\forall$ $b, b' \in B \setminus \{c(A)\}$ and B such that $B \setminus A = \emptyset$ and $c(A) \in B$.*

We will now show that the choice induced by f from any choice set is equal to the choice implied by c. First, note that this holds for X, since by Observation 1, $f(a, X, X, U^A) = f(b, X, X, U^A)$ for every $a, b \in X$ and every A with which there is an IIA violation associated. Moreover, $f(c(X), X, X, u^*) > f(a, X, X, u^*)$ $\forall$ $a \in X \setminus \{c(X)\}$ by P2. Then repeated application of P3 implies $f(c(X), X, X, U_c) > f(a, X, X, U_c)$ $\forall$ $a \in X \setminus \{c(X)\}$. Next, consider any $A \subsetneq X$ which causes an IIA violation. Suppose $A \in I_j$. Observation 1 implies that for any $B \in (\bigcup_{l=1}^{j} I_l) \setminus A$, $f(a, A, U^B) = f(a', A, U^B)$ $\forall$ $a, a' \in A$, and $f(c(A), A, X, U^A) > f(a, A, X, U^A)$ $\forall$ $a \in A$. Then repeated implication of P3 implies $f(c(A), A, X, U_j) > f(a, A, X, U_j)$ $\forall$ $a \in A$. By construction then $f(c(A), A, X, U_c) > f(a, A, X, U_c)$ $\forall$ $a \in A$. Finally, consider a set A that does not cause an IIA violation. One reason why this could happen is that condition (2) does not hold. This means $\mathcal{V}(A) \neq \emptyset$ and for every $B' \in \mathcal{V}(A)$, $c(B') = c(A)$. Let j^* be the smallest j such that for some $A' \in I_j$, $A \subset A'$ and $c(A') \in A$. Then the construction of U_c implies that $f(c(A), A, X, U_c) > f(a, A, X, U_c)$ $\forall$ $a \in A$, and the proof is complete. Therefore, assume that condition (2) holds but condition (1) does not hold. 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, $f(a, A, X, U^B) = f(a', A, X, U^B)$ $\forall$ $a, a' \in A$ and B with which an IIA violation is associated. Repeated use of P3, together with P2, implies $f(c(A), A, X, U_c) > f(a, A, X, U_c)$ $\forall$ $a \in A$.

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

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

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

$f(c(A), A, X, U^B) > f(a, A, X, U^B)\ \forall\ a \in A$, or $f(a, A, X, U^B) = f(a', A, X, U^B)\ \forall\ a, a' \in A$. But then repeated application of P3 implies that $f(c(A), A, X, U_j) > f(a, A, X, U_j)\ \forall\ a \in A$. By construction, $f(c(A), A, X, U_c) > f(a, A, X, U_c)\ \forall\ a \in A$.

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 and (2) holds in this case, 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 (2) holds for A'_j then A'_j causes an IIA violation with respect to the first such A''; if (2) does not hold for A'_j then B_j cannot be an IIA violation and the set of smallest supersets containing A'_j whose choice is in A'_j and causes a violation, all have choice equal to $c(A)$. If for every A'' it is the case that $c(A'') \notin A'_j$, then once again A'_j either (2) holds and A'_j causes an IIA violation with respect to B, or (2) does not hold, B_j cannot be an IIA violation, and the set of smallest supersets containing A'_j whose choice is in A'_j and causes a violation, all have choice equal to $c(A)$. Either way, we added members to ensure the choice $c(A)$ for every A'_j. This means that a should be the choice from A unless for some set B' between some A'_j and A we have $c(B') \in A \setminus \{a\}$ and members were added. But such a set cannot exist by minimality of the B_j's. ∎

Proof of Theorem 1

Let $X = \{a, b, c\}$ and take any $f \in \mathcal{F}^*$. For compactness, we use the notation $x_1 = f(a, \{a, b, c\}, X, U) - f(b, \{a, b, c\}, X, U)$, $x_2 = f(b, \{a, b, c\}, X, U) - f(c, \{a, b, c\}, X, U)$, $x_3 = f(a, \{a, c\}, X, U) - f(c, \{a, c\}, X, U)$, $x_4 = f(b, \{b, c\}, X, U) - f(c, \{b, c\}, X, U)$, and $x_5 = f(a, \{a, b\}, X, U) - f(b, \{a, b\}, X, U)$.

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

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

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

Then, determinants of three possible 5×5 matrices, each composed of five of the columns above, may be calculated to obtain:

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

To prove the result, it suffices to show that there exists U such that defining $x_1, x_2, \ldots, x_5$ 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 U corresponding to the i-th column is multiplied by c_i. The resulting group 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 $U \in \mathcal{U}(\{a, b, c\})$ such that $x_3 \neq x_4 + x_5$ and $f \circ U$ rationalizes choice where the worst element in the transitive pairwise ranking is best in the triple. Then either $2x_1 + x_2 \neq x_3 + x_5$ or $x_1 + 2x_2 \neq x_3 + x_4$.*[14]

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

Proof. By neutrality and symmetry of the condition $x_3 - x_4 - x_5 \neq 0$, there are two types of choice behaviors we must examine to prove the result:
Case 1: $a \succ_P b \succ_P c$ on the pairs, and $c \succ_T b \succeq_T a$ on the triple. That is, $x_3, x_4, x_5 > 0$, with $x_1 \leq 0$ and $x_2 < 0$. But then $2x_1 + x_2 \neq x_3 + x_5$, since the LHS is negative and the RHS is positive.
Case 2: $a \succ_P b \succ_P c$ on the pairs, and $c \succ_T a \succeq_T b$ on the triple. That is, $x_3, x_4, x_5 > 0$, with $x_1 \geq 0$, $x_2 < 0$. If we can find U such that $f \circ U$ rationalizes this behavior, then observe that $x_1 + 2x_2$ is negative. Hence $x_1 + 2x_2 \neq x_3 + x_4$ because the RHS is positive. ∎

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 $U = \{u\}$. 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 not degenerate 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 members in the n-dimensional simplex determines the aggregated utilities of n alternatives for all possible members over n alternatives, since any member is a scalar multiple of exactly one member 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 member's utilities get aggregated in pairs, and O_2 determines how a member'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 ease of exposition, let

$$\Gamma^l_1(f, v) = f(a, \{a, c\}, v) - f(c, \{a, c\}, v),$$

$$\Gamma^l_2(f, v) = f(a, \{a, b\}, v) - f(b, \{a, b\}, v) + f(b, \{b, c\}, v) - f(c, \{b, c\}, v),$$

$2x_1 + x_2 \neq x_3 + x_5$, as the RHS is positive.

$$\Gamma_2^l(f,v) = f(a,\{a,b,c\},v) - f(b,\{a,b,c\},v)] + [f(a,\{a,b,c\},v) - f(c,\{a,b,c\},v),$$

$$\Gamma_2^l(f,v) = [f(a,\{a,b\},v) - f(b,\{a,b\},v)] + [f(a,\{a,c\},v) - f(c,\{a,c\},v)],$$

for every $v \in \mathcal{F}^*$. Note that $\Gamma_i^j(v)$ stands for side j of the equation in condition i in the definition of a degenerate aggregator, given aggregator f and a member v.

1. Openness. Suppose that for aggregator f there is a member u over a triple such that neither of the equalities in the definition of a degenerate aggregator hold with equality. Note that u cannot be an indifferent member. Let $\varepsilon_i = \Gamma_i^l(f,v) - \Gamma_i^r(f,v)$ 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 u is not an indifferent member, $\alpha > 0$. Then for $\delta < \frac{\varepsilon}{8\alpha}$ it holds that $\Gamma_i^l(f',v) \neq \Gamma_i^r(f',v)$ 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 member $u \in \Delta_3$ 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} be such that $\alpha^{ij}(u(i),u(j)) \in \Delta^2$. Let $\alpha = \max(|\alpha^{ab}|, |\alpha^{ac}|, |\alpha^{bc}|)$. If for an aggregator f neither of the equalities in the definition of a degenerate aggregator hold, then the aggregator is by definition non-degenerate, hence there is trivially a point in the δ-neighborhood of $\beta(f)$ that corresponds to a non-degenerate aggregator. Otherwise let $\varepsilon \in (0, \frac{\delta}{\alpha})$ be such that $\varepsilon \neq |\Gamma_i^l(f,v) - \Gamma_i^r(f,v)|$ for $i \in \{1,2\}$.

Consider now 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 $v(x) - v(y) \geq u(a) - u(c)$. In words, with respect to members for which the utility difference between the elements of the pair is at least $u(a) - u(c)$ the aggregated utility is $\varepsilon > 0$ higher than what f yields for the preferred alternative (while it is the same for the other alternative) - otherwise f' is equivalent to f. By construction, $|\beta(f') - \beta(f)| < \delta$. Also note that $\Gamma_1^l(f',v) = \Gamma_1^l(f,v) + \varepsilon$, $\Gamma_1^r(f',v) = \Gamma_1^r(f,v)$, $\Gamma_2^l(f',v) = \Gamma_2^l(f,v)$, 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. ∎

Theorem 1 then follows from Theorem 2, Lemma 1, Lemma 2 and Lemma 3. ∎

References

Apesteguia, Jose and Miguel Ballester, "The Computational Complexity of Rationalizing Behavior," *Journal of Mathematical Economics*, 2010, pp. 356–363.

Arrow, Kenneth J. and Hervé Raynaud, *Social Choice and Multicriterion Decision-Making*, Cambridge, Massachussetts: The MIT Press, 1986.

Benabou, Roland and Marek Pycia, "Dynamic Inconsistency and Self-Control: A Planner-Doer Interpretation," *Economics Letters*, 2002, *77*, 419424.

Bernheim, Douglas and Antonio Rangel, "Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics," *Working Paper*, 2007.

Browning, Martin and Pierre-André Chiappori, "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests," *Econometrica*, 1998, *66*, 1241–1278.

Chambers, Chris and Takashi Hayashi, "Choice and Individual Welfare," *Working Paper*, 2008.

Chatterjee, Kalyan and R. Vijay Krishna, "Menu Choice, Environmental Cues and Temptation: A 'Dual Self' Approach to Self-Control," *American Economic Journal: Microeconomics*, forthcoming.

Cherepanov, Vadim, Tim Feddersen, and Alvaro Sandroni, "Rationalization," *Working paper*, 2010.

Chiappori, Pierre-André 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.

Dhillon, Amrita and Jean-Francois Mertens, "Relative Utilitarianism," *Econometrica*, 1999, *67*, 471–498.

Donahue, Eileen, Richard Robins, Brent Roberts, and Oliver John, "The Divided Self: Concurrent and Longitudinal Effects of Psychological Adjustment and Social Roles on Self-Concept Differentiation," *Journal of Personality and Social Psychology*, 1993, *64*, 834–846.

Evren, Ozgur and Efe Ok, “On the Multi-Utility Representation of Preference Relations,” *Working Paper*, 2007.

Fudenberg, Drew and David Levine, “A Dual Self Model of Impulse Control,” *American Economic Review*, 2006, *96*, 1449–1476.

Green, Jerry and Daniel Hojman, “Choice, Rationality, and Welfare Measurement,” *Working Paper*, 2009.

Kalai, Gil, Ariel Rubinstein, and Ran Spiegler, “Rationalizing Choice Functions By Multiple Rationales,” *Econometrica*, 2002, *70*, 24812488.

Kamenica, Emir, “Contextual Inference in Markets: On the Informational Content of Product Lines,” *American Economic Review*, 2008, *98*, 2127–2149.

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

Karni, Edi, “Impartiality: Defininition and Representation,” *Econometrica*, 1998, *66*, 1405–1415.

Keeney, Ralph L. and Howard Raiffa, *Decisions with Multiple Objectives*, Cambridge, U.K.: Cambridge University Press, 1993.

Kőszegi, Botond and Adam Szeidl, “A Model of Focusing in Economic Choice,” *Working Paper*, 2012.

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

Kochov, Asen, “The Epistemic Value of a Menu and Subjective States,” *Working Paper*, 2007.

Lachmann, Frank M., “How Many Selves Make a Person?,” *Contemporary Psychoanalysis*, 1996, *32*, 595–614.

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

Masatlioglu, Yusufcan and Daisuke Nakajima, “Theory of Choice By Elimination,” *Working Paper*, 2007.

____ **and Efe Ok**, “Rational Choice with Status Quo Bias,” *Journal of Economic Theory*, 2005, pp. 1–29.

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

Mitchell, Stephen, *Hope and Dread in Psychoanalysis*, New York: BasicBooks, 1993.

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

Power, M.J., "The Multistory Self: Why the Self is More Than the Sum of Its Autoparts," *Journal of Clinical Psychology*, 2007, *63*, 187–198.

Saari, Donald G., "Explaining All Three-Alternative Voting Outcomes," *Journal of Economic Theory*, 1999, *87*, 313–355.

Salant, Yuval 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., "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.

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 Non-anonymous aggregators

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

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

This extension allows us to consider asymmetric aggregators.

Example 1 (Asymmetric contextual concavity model)**.** *Interpret each member as corresponding to a product attribute, for which the preference belongs to a certain type. The class of preferences is parametrized by a concavity index. The contextual concavity aggregator in Kivetz et al. (2004) is given by*

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

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

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

Theorem 3. *Suppose f satisfies P1-P6 and is triple-solvable with k_f selves. Then, using n selves, f can rationalize any choice function c, defined on any finite grand set of alternatives X, that exhibits at most $\frac{n-1}{k_f}$ IIA violations.*

Consider a different type of example.

Example 2 (Costly self-control aggregators)**.** *Fudenberg and Levine (2006) propose a dual-self impulse control model with a long-run self exerting costly self-control over a short-run self. The reduced-form model they derive has an analogous representation in our framework, with two selves: the long-run self, with utility given by u^{lr} (the expected present value of the utility stream induced by the choice in the present), and the short-run self, with utility function u^{sr} (the present period consumption utility).*[15] *Using our terminology, the reduced form representation of their model assigns to alternative a the aggregate utility $u^{\mathrm{lr}}(a) - C(a)$, where term $C(a)$ depends on the attainable utility levels for the short-run self and is labeled as the cost of self-control. For example, using Fudenberg and Levine (2006)'s parametrization, $C(a) = \gamma[\max_{a' \in A} u^{\mathrm{sr}}(a') - u^{\mathrm{sr}}(a)]^{\psi}$.*

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

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

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

[15] 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.

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

Our results can be extended in a variety of ways to accommodate such restrictions. The most straightforward one imposes an assumption on the set $\mathcal{C}$ (which is satisfied in Example 2). Assume the existence of a type t and a collection of types $\widehat{T}$ such that appending any number of t-types to $\widehat{T}$ results in a collection of types in $\mathcal{C}$. In the generalized costly self-control aggregator above, the short-run type being t and the singleton set of a long-run type as $\widehat{T}$ satisfy this requirement. Let T_t^n denote the collection of n t-types. An aggregator f is *expandable* with $t \in T$ from $\widehat{T} \in \mathcal{C}$ if $(\widehat{T}, T_t^n) \in \mathcal{C}$ for every $n \in Z_+$. For an aggregator that is expandable with t from $\widehat{T}$ we can define triple-solvability with k type-t selves from $\widehat{T}$ as the existence of a collection of selves consisting of $|\widehat{T}|$ exactly indifferent selves over the triple whose type-composition is as in $\widehat{T}$ and k δ-indifferent selves of type t, such that the above collection of types constitutes a triple-basis for every $\delta > 0$.

Given the above definitions, the following result is obtained.

Theorem 4. *Suppose f is triple-solvable with k type-t selves from $\widehat{T}$. Then, using n selves, f can rationalize any choice function c, defined on any finite grand set of alternatives X, that exhibits at most $\frac{n-1-|\widehat{T}|}{k}$ IIA violations.*

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

B Examples rationalizing common choice procedures

Example 3 (The Median Procedure)**.** *The median procedure is a simple choice rule defined in Kalai et al. (2002). There is a strict ordering $\succ$ defined over elements of X, and the DM always chooses the median element of each $A \subseteq X$ according to $\succ$ (choosing the right-hand side element among the medians from choice sets with even number of alternatives).*

To rationalize this behavior, we consider the following aggregator.

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

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.*

Indeed, we claim that with the above 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 defined to be the product of the two terms.*

This rationalization is relatively simple and intuitive: the above selves are defined such that the DM is torn between two motivations, one in line with ordering $\succ$, and one going in exactly the opposite direction. Moreover, the geometric aggregation of these preferences drives the DM to choose the most central element of any choice set.

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 4 (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 $g : X \times P(X) \times \mathcal{X} \times R^X \to \mathbb{R}$ be such that*

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

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 $U = \{u_1, ..., u_n\} \in \mathcal{U}(X)$, let $f(a, A, X, U) = \sum\limits_{i=1}^{n} g(a, A, X, u_i)$. 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 one 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 behavior, 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. However, a given aggregator might act differently on a different collection of 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 possible selves).

C Approximate triple-solvability

For some aggregators a tighter upper bound can be given for the minimum group size needed to rationalize a choice function, by weakening the triple-solvability requirement. It suffices for triple-solvability to hold only *approximately*, which can yield a triple-basis with a smaller group size. For ease of exposition, we state this property for additively separable aggregators.

Definition 4. *We say $\hat{U} \in \mathcal{U}(\{a,b,c\})$ is a (δ,ε)-approximate triple-basis for f with respect to $\{a,b,c\}$ if $f(a,\{a,b\},\{a,b,c\},\hat{U}) = f(b,\{a,b\},\{a,b,c\},\hat{U}) + \delta$ and $|f(x,A,\{a,b,c\},\hat{U}) - f(y,A,\{a,b,c\},\hat{U})| < \varepsilon$ for all other $A \subseteq \{a,b,c\}$ and $x,y \in A$.*

That is, a group U 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 U is ε-indifferent among all alternatives given every other choice set.

We say that an aggregator f is *approximately triple-solvable with k members* if there is $\overline{\delta} > 0$ such that exists a (δ,ε)-approximate triple-basis with k members for every $\delta < \overline{\delta}$ and $\varepsilon > 0$. That is, for approximate triple-solvability we do not require that the group in 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. Theorem 2 can then be modified as follows.

Theorem 5. *Suppose f satisfies P1-P6 and P9, and is approximately triple-solvable with k_f members. 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 members.*

Proof. The only difference compared to the proof of Theorem 2 is in the construction of the rationalizing group. Recall the definition of $(I_j)_{j=1,...,j^*}$ 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 $\delta_j \in (0, \frac{\delta_{j-1}}{IIA(c)+1})$. Define a member u^X such that u^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 group $U^A \in \mathcal{U}(X)$ the following way: take a (δ_j, ε)-approximate triple-basis U, and define U^A by defining, for each $u_i \in U$, a member $u_i^A \in U^A$ by

$$u_i^A(x) = \begin{cases} u_i(a) & x = c(A) \\ u_i(b) & x \in A \setminus \{c(A)\} \\ u_i(c) & x \in X \setminus A. \end{cases}$$

Proving the group consisting of u^X and U^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 u defined on $\{a, b, c\}$, define the set $E(u, X) = \{\widehat{u} : X \to \{u(a), u(b), u(c)\} | \widehat{u}(x) = u(x)$ $\forall\ x \in \{a, b, c\}\}$. In words, $E(u, X)$ is the set of extensions of u from $\{a, b, c\}$ to X for which each element in $X/\{a, b, c\}$ receives the same utility as either a or b or c. Similarly, for any $U = (u_1, ..., u_m) \in \mathcal{U}(\{a, b, c\})$, let $E(U, X) = \{(\widehat{u}_1, ..., \widehat{u}_m) | \widehat{u}_i \in E(u_i, X)$ for all $i \in \{1, ..., m\}\}$.

Definition 5. *We say $U \in \mathcal{U}(\{a, b, c\})$ is a universal triple-basis for f if for any $X \supset \{a, b, c\}$ the following holds: for all $\hat{U} \in E(U, X)$, $f(a, \{a, b\}, X, \hat{U}) > f(b, \{a, b\}, X, \hat{U})$, and $f(\cdot, A, X, \hat{U})$ 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 members in the triple-basis. An aggregator f is *universally triple-solvable* if the following condition is satisfied.

Condition (Universal triple-solvability of f) There exists a triple $\{a, b, c\}$ and $k \in Z_+$ such that for every $\delta > 0$ there is a δ-indifferent $U \in \mathcal{U}^k(\{a, b, c\})$ constituting 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 members, 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.

Theorem 6. *Suppose f satisfies P1-P5 and is universally triple-solvable wrt to X with k_f members. Then, using n group members, f can rationalize any choice function, on any grand set of alternatives X, that exhibits at most $\frac{n-1}{k_f}$ IIA-violations.*