

A Behavioral Arrow Theorem

Alan D. Miller* and Shiran Rachmilevitch[†]

March 18, 2013

Abstract

In light of research indicating that individual behavior may violate standard rationality assumptions, we introduce a model of preference aggregation in which neither individual nor collective preferences must satisfy transitivity or other coherence conditions. We introduce an ordinal rationality measure to compare preference relations in terms of their level of coherence. Using this measure, we introduce a new axiom, monotonicity, which requires the collective preference to become more rational when the individual preferences become more rational. We show that no collective choice rule satisfies monotonicity and the standard Arrovian assumptions: unrestricted domain, weak Pareto, independence of irrelevant alternatives, and nondictatorship.

Keywords: Aggregation; Axioms; Intransitivity; Coherence; Monotonicity; Rationality; Arrow's Theorem.

JEL Codes: D60; D70; D71.

*Faculty of Law and Department of Economics, University of Haifa, Mount Carmel, Haifa, 31905, Israel. Email: admiller@econ.haifa.ac.il Web: <http://econ.haifa.ac.il/~admiller/>

[†]Department of Economics, University of Haifa, Mount Carmel, Haifa, 31905, Israel. Email: shiranrach@econ.haifa.ac.il Web: <http://econ.haifa.ac.il/~shiranrach/>

[‡]Acknowledgements will come in a final version.

1 Introduction

The core model of preference aggregation introduced by Arrow (1963) contains a strong assumption about the rationality of preferences. In particular, individual preferences are assumed to be reflexive, complete, and transitive. The first two properties are often considered richness conditions, while transitivity is a coherence condition. (See Bossert and Suzumura (2007).) In light of behavioral research casting doubt on the assumption of transitivity (Tversky, 1969), we modify Arrow’s model to remove the requirement that individual preference relations be transitive or satisfy other known coherence conditions.¹

Having removed this rationality requirement, we use this new framework to study an important question about the coherence of collective preferences. Do more rational individuals create a more rational society?

In other words, suppose that we have a set of agents, at least one of whom is less than fully rational, with a not-necessarily-rational collective preference. We may be able to induce an agent to “correct” his preferences by pointing out that his behavior violates minimal conditions of rationality. When the agent becomes more rational as a result of this correction, will the collective preference become more rational as a result?

Our conclusion is negative; if group decisions are made in a non-dictatorial way, it is possible that an increase in individual rationality may lead to a decrease in collective rationality. It may be possible to manipulate a group by helping individuals correct their mistakes. On a more practical note, a group of people may become more susceptible to “Dutch books” when the individuals’ susceptibility lessens.

We illustrate this problem by means of a simple example. It is clear that many collective choice rules satisfy the remaining assumptions imposed by Arrow (1963): unrestricted domain, weak Pareto, independence of irrelevant alternatives, and non-dictatorship. A simple example is the method of majority decision, in which alternative x is weakly preferred to alternative y whenever the majority weakly prefers x to y . (For more on the method of majority decision, see Sen (1964, 1966).)

However, the method of majority decision has an undesirable property. Suppose that there are three individuals, Alice, who prefers x to y to z , Bob, who prefers z to x to y , and Carol who prefers x to z to y to x . Alice and Bob have transitive preferences, but Carol’s preferences are not. By the method of majority decision, x is preferred to y , z is preferred to y , and x is preferred to z , leading to a transitive and rational collective preference. However, suppose that Carol realizes that her preferences are irrational and seeks to “correct” them. She decides to retain her view that y is preferred to x but changes her opinion of z , so that she now prefers y to z and z to x . As a consequence, the method of majority decision leads to the collective preference x to y to z to x , and is no longer transitive. In this case, the collective

¹With a similar motivation, de Clippel (2012) studies mechanism design without the assumption that individuals make choices as if they are maximizing a preference relation.

preference became less rational *because* Carol became more rational.

We show that this problem is not unique to method of majority decision. In fact, every collective choice rule which satisfies the remaining assumptions of Arrow (1963) will have this undesirable property. For every such method it is possible that an increase in the coherence of individual preferences will lead to a decrease in the coherence of the collective preference.²

Our formal model can be described as follows. First, we study a modified version of collective choice rules in which neither the individual nor collective preferences are required to be rational. We assume only that preference relations be reflexive and complete. Thus each individual's preferences can be described by a reflexive and complete relation, and the the collective preference can be described by a reflexive and complete relation as well. We implicitly assume that every possible combination of individual preference relations is possible; i.e. that Arrow's unrestricted domain axiom holds in this setting.

Using accepted notions of rationality, we introduce the concept of an ordinal rationality measure, and identify some minimal conditions that any reasonable rationality measure should satisfy. We formulate an axiom, *monotonicity*, to address the problem exhibited by majority rule in the above example. This axiom requires that the collective choice rule be monotonic with respect the rationality measure; that is, if individual preferences change and become more rational, then the collective preference should become more rational, if it changes at all.

In addition to monotonicity, we impose the three additional axioms of Arrow (1963): *weak Pareto*, which requires that society strictly prefer x to y whenever every individual strictly prefers x to y , *independence of irrelevant alternatives*, which requires that a change in the opinions about alternative z does not affect the relative ranking of alternatives x and y , and *nondictatorship*, which requires that no individual be a dictator. We show that the four axioms are incompatible. In other words, regardless of which ordinal rationality measure we choose, we cannot find a collective choice rule which is monotonic, weakly Paretian, independent of irrelevant alternatives, and nondictatorial.

1.1 Relevant Literature

Previous studies have sought to weaken the assumption of rationality in Arrow (1963) by permitting a wider range of collective preferences. The case of quasi-transitive collective preferences was studied by many including Gibbard (1969), Sen (1969, 1970), Schick (1969), and Mas-Colell and Sonnenschein (1972), and of acyclic preferences by

²For this reason, we do not argue that the method of majority decision is any worse than any other method in this context. Several studies, including Sen (1966), Inada (1969), and Batra and Pattanaik (1972), examine the conditions under which pairwise majority does not lead to cycles. Dasgupta and Maskin (2008) provide an argument that the method of majority decision is more robust than other voting methods in that it violates the standard axioms on fewer domains.

Mas-Colell and Sonnenschein (1972) and Blair and Pollak (1982). For more see Sen (1977).³

Other scholars have tried to avoid the negative conclusions of Arrow (1963) by moving in the opposite direction. Instead of expanding the range of admissible collective preferences, these studies restrict the domain of allowable preferences. The most prominent example is that of the single-peaked preference restriction of Black (1948a,b) and Arrow (1963).

The most closely related literature is the study of tournaments, which are described by binary relations which are antisymmetric and complete. Unlike the preference relations we study, tournaments do not allow for the possibility of ties. In this context, Monjardet (1978) shows that a collective choice rule that (a) maps every profile of transitive preferences into a transitive preference, (b) satisfies the independence of irrelevant alternatives axiom and (c) satisfies a non-imposition axiom is either dictatorial or “persecutive.” Roughly speaking, persecutive means that the decisive coalitions are all the coalitions that do not contain a certain individual i . A related result can be found in Barthelemy (1982). As far as we can tell, the monotonicity axiom that we present is new to this paper.

The violation of transitivity is one of the simplest and most basic violations of rationality. There are, of course, more sophisticated violations. Rubinstein and Salant (forthcoming), for example, consider a decision maker whose behavior is consistent with the maximization of one (transitive) preference relation under some circumstances, and consistent with the maximization of another (transitive) preference relation under other circumstances. Related works by Manzini and Mariotti (2007), Cherepanov et al. (2008), and Masatlioglu et al. (forthcoming) consider a decision maker who first identifies a subset of alternatives from the grand set of all alternatives, and then maximizes a transitive relation on this subset. We do not identify rationality and transitivity; we only claim that any reasonable definition of rationality will contain a transitivity requirement.

There are concepts of rationality which are inconsistent with our approach. First among these is are the postulates of decision theory, such as those of Savage (1972), which some may identify with the concept of rationality. While transitivity is among the axioms, we do not have a method to evaluate tradeoffs between a violation of transitivity and, for example, a violation of the sure thing principle. Second, there are some preferences which satisfy classical notions of rationality and which nonetheless seem bizarre; for example, complete indifference. One might view a person who is completely indifferent between all alternatives as being less rational than an individual whose preferences are mostly consistent with the exception of a single cycle among three alternatives. The relevance of this latter case to our analysis is somewhat attenuated by the fact that we do not need the full strength of transitivity for our results, but only a very weak condition described in Section 3.1.

³A related literature looks at the case in which preferences are transitive but not necessarily complete. (See Baucells and Shapley, 2008; Pini et al., 2009)

2 Model and result

Let X be a set of alternatives, $|X| \geq 3$. A preference relation R on X is (a) **complete** if for all $x, y \in X$, $x \neq y$ implies that either xRy or yRx , (b) **reflexive** if for all $x \in X$, xRx , and (c) **transitive** if for all $x, y, z \in X$, xRy and yRz implies that xRz . Let \mathcal{R} be the set of all complete and reflexive preference relations on X . A **preference ordering** is a preference relation which is complete, reflexive, and transitive. Let $\mathcal{R}^* \subseteq \mathcal{R}$ be the set of preference orderings on X . For a preference relation $R \in \mathcal{R}$ we denote by P its asymmetric component; that is, xPy if xRy but not yRx .

For $Y \subseteq X$, denote by $\mathcal{R}|_Y$ the set all complete and reflexive preference relations on Y , and denote by $\mathcal{R}^*|_Y$ the set all preference orderings in $\mathcal{R}|_Y$. For $R \in \mathcal{R}$ and $Y \subseteq X$, denote by $R|_Y \in \mathcal{R}|_Y$ the restriction of R to Y .

Let $N \equiv \{1, \dots, n\}$ be a finite set of agents, $n \geq 2$. A *profile* $R = (R_1, \dots, R_n) \in \mathcal{R}^N$ is a vector of binary relations, one for each agent. A **collective choice rule** is a mapping $f: \mathcal{R}^N \rightarrow \mathcal{R}$.⁴ We define $R_0 \equiv f(R)$ to be the social relation, and we denote by P_0 its asymmetric component.

A **rationality measure** is a binary relation \succsim on \mathcal{R} which satisfies the following properties:

1. For all $R \in \mathcal{R}$, $R \succsim R$.
2. For all $R' \in \mathcal{R}^*$ and $R \in \mathcal{R}$, $R \succsim R'$ implies that $R \in \mathcal{R}^*$.
3. For all $R' \in \mathcal{R}^*$ and $R \in \mathcal{R} \setminus \mathcal{R}^*$, $R' \succsim R$.

For two profiles $R, R' \in \mathcal{R}^N$ we write $R \succsim R'$ if $R_i \succsim R'_i$ for all $i \in N$.

Property 1, known as reflexivity, requires each preference relation to be “at least as rational” as itself. Property 2 requires that only a preference ordering can be at least as rational as another preference ordering. Property 3 requires that every preference ordering be at least as rational as every non-transitive preference relation.⁵

A wide range of rationality measures satisfies these conditions. We provide two examples. The simplest rationality measure \succsim' is one for which $R^* \succsim' R$ if and only if $R^* \in \mathcal{R}^*$ and $R \in \mathcal{R} \setminus \mathcal{R}^*$. A more complicated rationality measure can incorporate the structure of coherence properties studied in the social choice literature. A preference relation is **quasi-transitive** if, for all $x, y, z \in X$, xPy and yPz together imply that xPz , and is **acyclic** if, for every $k \geq 3$ and every $x^1, \dots, x^k \in X$, x^iPx^{i+1} for all $i < k$ implies that x^kPx^1 does not hold. Define $\mathcal{R}^q \subseteq \mathcal{R}$ as the set of preference relations which are complete, reflexive, and quasi-transitive, and define $\mathcal{R}^a \subseteq \mathcal{R}$ as the set of preference relations which are complete, reflexive, and acyclic. It is well known

⁴This is a slight change from the standard definition, in which the domain of a collective choice rule is a set of preference orderings (see Sen, 1970).

⁵Properties 2 and 3 can be weakened without changing our results. This is discussed Section 3.2, below.

that $\mathcal{R}^* \subsetneq \mathcal{R}^a \subsetneq \mathcal{R}^b \subsetneq \mathcal{R}$ (see Suzumura, 1983). Thus, we can define a rationality measure \succ'' such that $R^* \succ'' R$ if and only if there exists an $\mathcal{C} \in \{\mathcal{R}^*, \mathcal{R}^a, \mathcal{R}^b\}$ such that $R^* \in \mathcal{C}$ but $R \notin \mathcal{C}$.⁶

Our first axiom, monotonicity, requires that if preference relations change, and each individual's new preference relation stays at least as rational as it was before the change, then the social preference must stay at least as rational.

Monotonicity: For all $R, R' \in \mathcal{R}^N$, if $R \succ R'$ then $R_0 \succ R'_0$.

The following three axioms were introduced by Arrow (1963); for brevity, we will not discuss them.

Weak Pareto: For every $R \in \mathcal{R}^N$ and $x, y \in X$, if xP_iy for all $i \in N$, then xP_0y .

Independence of Irrelevant Alternatives: For all $Y \subseteq X$ and $R, R' \in \mathcal{R}^N$, if $R|_Y = R'|_Y$, then $R_0|_Y = R'_0|_Y$

An individual $d \in N$ is a *dictator* if, for all $R \in \mathcal{R}^N$, xP_dy implies that xP_0y .

Non-Dictatorship: There does not exist a dictator.

We can now turn to the main result.

Theorem 1. *There does not exist a collective choice rule that satisfies monotonicity, weak Pareto, independence of irrelevant alternatives, and non-dictatorship.*

To prove this Theorem we make use of the following lemma. For a coalition $K \subseteq N$, we define $x\bar{D}_Ky$ as the statement that the coalition K is decisive for x over y ; that is, if xP_iy for all $i \in K$, then xP_0y . Similarly, we define xD_Ky as the statement that the coalition K is decisive for x over y when all others are opposed; that is, if xP_iy for all $i \in K$ and yP_ix for all $i \notin K$, then xP_0y .

Lemma 1. *If a collective choice rule f satisfies monotonicity, weak Pareto, and independence of irrelevant alternatives, then whenever xD_Ky for a coalition $K \subseteq N$ and some pair of alternatives $x, y \in X$, it follows that $w\bar{D}_Kz$ for every pair $w, z \in X$.*

The proofs of the theorem and the lemma are given in the appendix.

⁶The four classes were chosen for the ease of the exposition, clearly a rationality measure can incorporate any number of classes, and these not be totally ordered through set inclusion. In particular, the rationality measure can incorporate the coherence properties of semi-transitivity and the interval order. See Cato (2011).

3 Discussion

3.1 Generalized Choice

We have represented preferences by means of binary relations. A more general approach would be to study choice functions. Formally, a **domain** $\mathcal{S} \subseteq 2^X \setminus \emptyset$ is a set of non-empty subsets of X . A **choice function** C on a **choice space** (X, \mathcal{S}) is a function which maps each element $S \in \mathcal{S}$ to a non-empty subset of X such that $C(S) \subseteq S$.

For a preference relation R define the choice function C^R such that, for all $S \in \mathcal{S}$, $C^R(S) \equiv \{x \in S : xRy \text{ for all } y \in S\}$. A choice function C is **rationalizable** if there exists a preference relation R such that $C = C^R$. In such a case, the preference relation R is said to **rationalize** C .

On the binary domain $\mathcal{S}_B \equiv \{\{x, y\} \subseteq X : x \neq y\}$, which consists of all two-element subsets of X , it is straightforward to see that a preference relation rationalizes a choice function if and only if it is complete and reflexive, and that all choice functions are rationalizable. Furthermore, there is a one-to-one mapping between the set of choice functions and the set of complete and reflexive preference relations. Consequently, the model of choice functions on the choice space (X, \mathcal{S}_B) is formally equivalent to that of the preference relation model studied in the previous section.

An interesting extension would be to consider the domain \mathcal{S}_F of finite subsets of X . Let \mathcal{C} be the set of all choice functions on (X, \mathcal{S}_F) . It is well known that a preference relation R rationalizes a choice function $C \in \mathcal{C}$ if and only if it is complete, reflexive, and acyclic. (Suzumura, 1983, Theorem 2.9) A choice function is **fully rational** if it has a rationalization which is reflexive, complete, and transitive. Let $\mathcal{C}^* \subseteq \mathcal{C}$ be the set of fully rational choice functions.

For $Y \subseteq X$ and a domain \mathcal{S} , let $\mathcal{S}|_Y \equiv \{S \in \mathcal{S} | S \subseteq Y\}$ be the restriction of \mathcal{S} to Y . Denote by $\mathcal{C}|_Y$ the set all choice functions on $(X, \mathcal{S}|_Y)$. For a choice function $C \in \mathcal{C}$ and $Y \subseteq X$, denote by $C|_Y \in \mathcal{C}|_Y$ the restriction of C to $(X, \mathcal{S}|_Y)$.

The rest of the definitions in this subsection will closely parallel those in Section 2. A **profile** $C = (C_1, \dots, C_n) \in \mathcal{C}^N$ is a vector of choice functions, one for each agent. An **collective choice function** is a mapping $f: \mathcal{C}^N \rightarrow \mathcal{C}$. We use the notation $C_0 \equiv f(C)$ to denote the collective choice function.

A **rationality measure** is a binary relation \succsim on \mathcal{C} which satisfies the following properties:

- 1a. For all $C \in \mathcal{C}$, $C \succsim C$.
- 2a. For all $C^* \in \mathcal{C}^*$ and $C \in \mathcal{C}$, $C \succsim C^*$ implies that $C \in \mathcal{C}^*$.
- 3a. For all $C^* \in \mathcal{C}^*$ and $C \in \mathcal{C} \setminus \mathcal{C}^*$, $C^* \succsim C$.

For two profiles $C, C' \in \mathcal{C}^N$ we write $C \succsim C'$ if $C_i \succsim C'_i$ for all $i \in N$.

Monotonicity: For all $C, C' \in \mathcal{C}^N$, if $C \succ C'$ then $C_0 \succ C'_0$.

Weak Pareto: For every $C \in \mathcal{C}^N$ and $x, y \in X$, if $C_i(\{x, y\}) = \{x\}$ for all $i \in N$, then $C_0(\{x, y\}) = \{x\}$.

Independence of Irrelevant Alternatives: For all $Y \subseteq X$ and $C, C' \in \mathcal{C}^N$, if $C|_Y = C'|_Y$, then $C_0|_Y = C'_0|_Y$.

An individual $d \in N$ is a *dictator* if, for all $C \in \mathcal{C}^N$ and $x, y \in X$, $C_d(\{x, y\}) = \{x\}$ implies that $C_0(\{x, y\}) = \{x\}$.

Non-Dictatorship: There does not exist a dictator.

Because the non-dictatorship axiom is defined with respect to binary choices, and because the independence of irrelevant alternatives axiom allows us to focus on binary choices, the conclusions we draw in this section are parallel to those of Theorem 1. For this reason, we state the following corollary without proof.

Corollary 1. *There does not exist a collective choice function f that satisfies monotonicity, weak Pareto, independence of irrelevant alternatives, and non-dictatorship.*

3.2 Weakening the axioms.

The monotonicity axiom can be weakened to allow for a broader class of rationality measures. In particular, properties 2 and 3 can be weakened to properties 2' and 3'. Recall that \mathcal{R}^a is the set of all acyclic preference relations; that is, those which do not contain P-cycles. A set of elements $Y \subseteq X$ is *top-ranked* in profile R if $a \in Y$, $b \in X \setminus Y$, and $i \in N$ implies that $aP_i b$.

2'. For all $R^* \in \mathcal{R}^*$ and $R \in \mathcal{R}$, $R \succ R^*$ implies that $R \in \mathcal{R}^a$.

3'. For all $R^* \in \mathcal{R}^*$ and $R \in \mathcal{R} \setminus \mathcal{R}^a$: if there is a three-element set $Y \subseteq X$ which is top-ranked in both relations, such that $R^*|_{X \setminus Y} = R|_{X \setminus Y}$, then $R^* \succ R$.

Property 2' requires that only an acyclic preference relation can be at least as rational as a transitive preference ordering. Because every transitive preference relation is also acyclic, this property is weaker than property 2. Property 3' changes property 3 in two ways. First, it applies only to comparisons between transitive preference orderings and cyclic preference relations. Second, it is limited to the specific case in which the cyclic preference relation and the transitive preference ordering are identical except for the three top-ranked elements. This is a very clear case in which the transitive relation is more rational than the cyclic one.

In the context of generalized choice functions, properties 2a and 3a can be weakened to properties 2a' and 3a'. Let $\mathcal{C}^R \subset \mathcal{C}$ be the set of rationalizable choice functions on the space (X, \mathcal{S}_F) .

2a'. For all $C^* \in \mathcal{C}^*$ and $C \in \mathcal{C}$, $C \succ C^*$ implies that $C \in \mathcal{C}^R$.

3a'. For all $C^* \in \mathcal{C}^*$ and $C \in \mathcal{C} \setminus \mathcal{C}^R$, $C^* \succ C$.

Property 2a' requires that only a rationalizable choice function can be at least as rational as a fully rational choice function. Because a choice function in \mathcal{C} is rationalizable if and only if it has a rationalization $R \in \mathcal{R}^a$, this property is a direct analogue of Property 2'. Property 3a' requires that every fully rational choice function must be at least as rational as every non-rationalizable choice function. This property is similarly analogous to property 3'. However, it is not limited to the case in which the cyclic preference relation and the transitive preference ordering are identical except for the three top-ranked elements. It is not clear how to interpret this additional restriction in the choice function setting.

By weakening the properties of the rationality measure, we consequently weaken the monotonicity axiom. This would not, however, change the main result. There does not exist a collective choice rule (or collective choice function) f that satisfies monotonicity with respect to any rationality measure satisfying 1, 2' and 3' (or 1, 2a', and 3a'), weak Pareto, independence of irrelevant alternatives, and non-dictatorship.

Alternatively, it is possible to replace transitivity with acyclicity entirely by using the following four properties. Note that each of these properties is weaker than the original property it replaces, but that each is stronger than its corresponding property above.

2''. For all $R^* \in \mathcal{R}^a$ and $R \in \mathcal{R}$, $R \succ R^*$ implies that $R \in \mathcal{R}^a$.

3''. For all $R^* \in \mathcal{R}^a$ and $R \in \mathcal{R} \setminus \mathcal{R}^a$: if there is a three-element set $Y \subseteq X$ which is top-ranked in both relations, such that $R^*|_{X \setminus Y} = R|_{X \setminus Y}$, then $R^* \succ R$.

2a''. For all $C^* \in \mathcal{C}^a$ and $C \in \mathcal{C}$, $C \succ C^*$ implies that $C \in \mathcal{C}^R$.

3a''. For all $C^* \in \mathcal{C}^a$ and $C \in \mathcal{C} \setminus \mathcal{C}^R$, $C^* \succ C$.

3.3 Relationship to Transitivity and Acyclicity

A natural question involves the extent to which the monotonicity axiom introduced in this paper substitutes for the standard assumption of transitivity or for other coherence conditions.

For example, consider the following axiom:

Transitive-to-Transitive: For all $R \in \mathcal{R}^{*N}$, $R_0 \in \mathcal{R}^*$.⁷

⁷We thank Eric Maskin for suggesting this idea.

The transitive-to-transitive axiom requires that every profile of transitive preference relations must map to a transitive social relation. There is no logical relation between this axiom and the monotonicity axiom we propose. For example, a constant rule that maps all profiles to the same non-transitive social preference satisfies monotonicity but not this axiom. To see that a rule may satisfy the transitive-to-transitive axiom but not monotonicity, consider a rule in which the social preference coincides with that of the first agent when that agent, and only that agent, has a non-transitive preference relation, and which otherwise maps to a fixed transitive social preference. When all agents' preferences are transitive, this rule will lead to a transitive social preference, and thus it satisfies the transitive-to-transitive axiom. If the first agent's preferences change and become non-transitive, the social preference will clearly become non-transitive. However, if a second agent's preferences change and become non-transitive, the social preference will change back to the original transitive preference, thus violating monotonicity.

However, in the presence of weak Pareto, independence of irrelevant alternatives, and non-dictatorship, the two axioms are equivalent. To see this, note that in the context of Arrow (1963), the transitive-to-transitive axiom implies that Arrow's condition of unrestricted domain is satisfied on the set of transitive profiles. Consequently, when combined with weak Pareto and independence of irrelevant alternatives, this axiom implies the existence of an individual d who is decisive over every pair of alternatives for every transitive profile. That is $xP_d y$ implies $xP_0 y$ for every transitive profile. By the independence of irrelevant alternatives axiom, however, it becomes irrelevant whether the profile is transitive; and hence individual d is a dictator. Thus we can derive the following proposition:

Proposition 1. *There does not exist a collective choice rule that satisfies transitive-to-transitive, weak Pareto, independence of irrelevant alternatives, and non-dictatorship.*

Because transitivity can be replaced with acyclicity without affecting the theorem, we might consider a parallel axiom which requires every profile of acyclic preference relations to map to an acyclic social relation.

Acyclic-to-Acyclic: For all $R \in \mathcal{R}^{aN}$, $R_0 \in \mathcal{R}^a$.

As with the transitive-to-transitive axiom, the acyclic-to-acyclic axiom bears no logical relation to our monotonicity axiom. This can be seen through a minor modification of the examples above, in which 'transitivity' is replaced with 'acyclicity'. Unlike the transitive-to-transitive axiom, however, acyclic-to-acyclic is not equivalent to monotonicity in the presence of the other three axioms. For example, consider the Pareto extension rule (Sen, 1969, 1970) which declares an alternative x to be strictly preferred to an alternative y if and only if every individual strictly prefers x to y , and which declares x to be indifferent to y when neither alternative is strictly preferred to the other. The Pareto extension rule satisfies the acyclic-to-acyclic, independent

of irrelevant alternatives, weak Pareto, and non-dictatorship axioms, but does not satisfy the monotonicity axiom (even if ‘transitivity’ is replaced with ‘acyclicity’ in the definition of the rationality measure.)

4 Conclusion

This paper departs from the standard approach to preference aggregation in three ways. First, in light of research indicating that individual behavior may violate standard assumptions of rationality, we modify the standard model of preference aggregation to study the case in which neither individual nor collective preferences are required to satisfy transitivity or other coherence conditions. Second, we introduce the concept of an ordinal rationality measure which can be used to compare preference relations in terms of their level of coherence. Third, using this measure, we introduce a monotonicity axiom which requires that the collective preference become more rational when the individual preferences become more rational. We show that for any ordinal rationality measure, it is impossible to find a collective choice rule which satisfies the monotonicity axiom and the other standard assumptions introduced by Arrow (1963): unrestricted domain, weak Pareto, independence of irrelevant alternatives, and nondictatorship.

Appendix

Proof of Lemma 1. We assume that the rationality measure satisfies properties 1, 2', and 3'. This will be sufficient to prove the lemma.

Let the collective choice rule f satisfy the monotonicity, weak Pareto, and independence of irrelevant alternatives axioms. Let $K \subseteq N$ and $x, y \in X$ such that $x D_K y$.

Step one. We claim that, for all $z \in X \setminus \{x, y\}$, if $R \in \mathcal{R}^N$ such that (a) $R_i|_{\{x, y, z\}} \in \mathcal{R}^*|_{\{x, y, z\}}$ for all $i \in N$, (b) $x P_i y$ for all $i \in K$, and (c) $R_i|_{\{x, y, z\}} = R_j|_{\{x, y, z\}}$ for all $i, j \in K$, then $R_0|_{\{x, y, z\}} \in \mathcal{R}^*|_{\{x, y, z\}}$. To prove this claim, let $z \in X \setminus \{x, y\}$ and let $R \in \mathcal{R}^N$ satisfying (a), (b), and (c). From the independence of irrelevant alternatives axiom we can assume, without loss of generality, that the set $\{x, y, z\}$ is top-ranked in each R_i . Let $R^\circ \in \mathcal{R}^N$ such that (i) $R_i^\circ = R_i$ for all $i \in K$, (ii) $y P_i^\circ x$, $x P_i^\circ z$, and $z P_i^\circ y$ for all $i \notin K$, and (iii) $R_i \succ R_i^\circ$ for all $i \in N$. Because $x D_K y$ it follows that $x P_0^\circ y$.

From condition (c) it follows that there are two cases: either $x P_i^\circ z$ for all $i \in K$, or $z R_i^\circ x$ for all $i \in K$. In the former case, $x P_i^\circ z$ for all $i \in N$, which implies (by weak Pareto), that $x P_0^\circ z$. Because $x P_0^\circ y$ and $x P_0^\circ z$, it follows that $R_0^\circ|_{\{x, y, z\}} \in \mathcal{R}^*|_{\{x, y, z\}}$. In the latter case, $z P_i^\circ y$ for all $i \in N$, which implies (by weak Pareto), that $z P_0^\circ y$. Because $x P_0^\circ y$ and $z P_0^\circ y$, it follows that $R_0^\circ|_{\{x, y, z\}} \in \mathcal{R}^*|_{\{x, y, z\}}$. Because $R_0^\circ|_{\{x, y, z\}} \in$

$\mathcal{R}^* \upharpoonright_{\{x,y,z\}}$ it follows from monotonicity and independence of irrelevant alternatives that $R_0 \upharpoonright_{\{x,y,z\}} \in \mathcal{R}^* \upharpoonright_{\{x,y,z\}}$, proving the claim.

Step two. Let $R' \in \mathcal{R}^{*N}$ such that, for all $i \in K$, xP'_iy and yP'_iz and, for all $i \notin K$, yP'_ix and yP'_iz . Because xD_ky it follows that xP'_0y , and because yP'_iz for all $i \in N$ it follows from weak Pareto that yP'_0z . Because R' satisfies requirements (a), (b), and (c) of step one, it follows that $R'_0 \upharpoonright_{\{x,y,z\}} \in \mathcal{R}^* \upharpoonright_{\{x,y,z\}}$ and therefore xP'_0z . By the independence of irrelevant alternatives axiom, this implies that $x\bar{D}_Kz$. In other words:

$$xD_Ky \text{ implies that } x\bar{D}_Kz. \quad (1)$$

Now, let $R'' \in \mathcal{R}^{*N}$ such that, for all $i \in K$, zP''_ix and xP''_iy and, for all $i \notin K$, zP''_ix and yP''_ix . Because xD_ky it follows that xP''_0y , and because zP''_ix for all $i \in N$ it follows from weak Pareto that zP''_0x . Because R'' satisfies requirements (a), (b), and (c) of step one, it follows that $R''_0 \upharpoonright_{\{x,y,z\}} \in \mathcal{R}^* \upharpoonright_{\{x,y,z\}}$ and therefore zP''_0y . By the independence of irrelevant alternatives axiom, this implies that $z\bar{D}_Ky$. In other words:

$$xD_Ky \text{ implies that } z\bar{D}_Ky. \quad (2)$$

By interchanging y and z in statement (2) it follows that:

$$xD_Kz \text{ implies that } y\bar{D}_Kz. \quad (3)$$

By replacing x by y , y by z , and z by x in statement (1) it follows that:

$$yD_Kz \text{ implies that } y\bar{D}_Kx. \quad (4)$$

By combining statements (1), (3), and (4) it follows that:

$$xD_Ky \text{ implies that } y\bar{D}_Kx. \quad (5)$$

By interchanging x and y in statements (1), (2), and (5) it follows that

$$\begin{aligned} yD_Kx \text{ implies that } y\bar{D}_Kz, \\ yD_Kx \text{ implies that } z\bar{D}_Kx, \\ yD_Kx \text{ implies that } x\bar{D}_Ky, \end{aligned}$$

and therefore by combining statement (5) it follows that:

$$xD_Ky \text{ implies that } y\bar{D}_Kz, z\bar{D}_Kx, \text{ and } x\bar{D}_Ky. \quad (6)$$

Therefore, we are led to the implication that:

$$\text{for every } \{x, y, z\} \subseteq X, \text{ if } xD_Ky \text{ then } a\bar{D}_Kb \text{ for every } a, b \in \{x, y, z\}. \quad (7)$$

Clearly, statement (7) applies if we replace z with w . By replacing z with w in

statement (1) it follows that

$$xD_Ky \text{ implies that } x\bar{D}_Kw. \quad (8)$$

By replacing y with w in statement (2) it follows that

$$xD_Kw \text{ implies that } z\bar{D}_Kw. \quad (9)$$

By replacing x with z and y with w in statement (5) it follows that

$$zD_Kw \text{ implies that } w\bar{D}_Kz. \quad (10)$$

By combining statements (7), (8), (9), and (10), we are led to the result that, for every $\{x, y\}, \{w, z\} \subseteq X$, if xD_Ky then $w\bar{D}_Kz$. This concludes the proof. \square

Proof of Theorem 1. We assume that the rationality measure satisfies properties 1, 2', and 3'. This will be sufficient to prove the theorem.

Let f be a collective choice rule that satisfies the monotonicity, weak Pareto, independence of irrelevant alternatives, and non-dictatorship axioms. We will derive a contradiction.

Let $S \subseteq N$ be coalition of minimal size, so that $|T| < |S|$ implies that xD_Ty is false for all $x, y \in X$. By the weak Pareto axiom, such a coalition S exists. By the non-dictatorship axiom and Lemma 1, $|S| \geq 2$. Without loss of generality, let xD_Sy . Let $S_1 \subseteq S$ such that $|S_1| = 1$, let $S_2 \equiv S \setminus S_1$, and let $S_3 \equiv N \setminus S$.

Let $R \in \mathcal{R}^{*N}$ be a transitive profile such that (a) xP_iy and yP_iz for all $i \in S_1$, (b) zP_ix and xP_iy for all $i \in S_2$, and (c) yP_iz and zP_ix for all $i \in S_3$. Let $R_* \in \mathcal{R}$ such that xP_*y , yP_*z , and zP_*x , and let $R_+ \in \mathcal{R}$ such that xP_+z , zP_+y , and yP_+x . Let $R^A, R^B, R^C \in \mathcal{R}^N$ be profiles such that (a) $R_i^A = R_i^B = R_i^C = R_*$ for all $i \in S_1$, (b) $R_i^A = R_i^B = R_i^C = R_+$ for all $i \in S_2$, and (c) $R_i^A = R_*$, $R_i^B = R_+$, and $R_i^C = R_i$ for all $i \in S_3$.

Because of the independence of irrelevant alternatives axiom, we can assume, without loss of generality, that the elements $x, y, z \in X$ are top-ranked in profiles R, R^A, R^B , and R^C and that $R|_{X \setminus \{x, y, z\}} = R^A|_{X \setminus \{x, y, z\}} = R^B|_{X \setminus \{x, y, z\}} = R^C|_{X \setminus \{x, y, z\}}$. It follows that $R \succcurlyeq R^A, R \succcurlyeq R^B$, and $R \succcurlyeq R^C$. Therefore, by monotonicity, if one or more of R_0^A, R_0^B , and R_0^C is transitive, then R_0 must be transitive.

Suppose, contrariwise, that R_0 is not transitive. It follows that neither R_0^A, R_0^B , nor R_0^C may be transitive. Because S_2 is not a decisive coalition, it follows that $xR_0^A y, yR_0^A z$, and $zR_0^A x$. Because R_0^A is not transitive it follows that $S_1 \cup S_3$ must be decisive for at least one of the three pairs x over y, y over z , or z over x . By Lemma 1, it follows that $xD_{S_1 \cup S_3} y$ for all $x, y \in X$.

Because S_1 is not a decisive coalition, it follows that $xR_0^B z, zR_0^B y$, and $yR_0^B x$. Because R_0^B is not transitive it follows that $S_2 \cup S_3$ must be decisive for at least one of the three pairs x over z, z over y , or y over x . By Lemma 1, it follows that $xD_{S_2 \cup S_3} y$ for all $x, y \in X$.

Because $x D_{S_1 \cup S_3} y$ for all $x, y \in X$ it follows that $y P_0^C z, z P_0^C x$. Because $x D_{S_2 \cup S_3} y$ for all $x, y \in X$ it follows that $y P_0^C x$. Therefore it follows that R_0^C is transitive, which is a contradiction, proving that R_0 must be transitive.

By assumption, the coalition $S = S_1 \cup S_2$ is decisive for x over y . This implies that $x P_0 y$. Because $z P_i y$ only for $i \in S_2$ and S_2 is not decisive, it follows that $y R_0 z$. Because R_0 is transitive, it follows that $x P_0 z$. But this means that $x D_{S_1} z$, which implies, by Lemma 1, that S_1 is a dictator. This violates the non-dictatorship axiom, and concludes the impossibility proof.

Independence of the Axioms. We describe four collective choice rules. Each of the rules satisfies three of the axioms while violating the fourth. This is sufficient to prove the independence of the axioms.

Rule 1. For all $x, y \in X$, let $x R_0 y$ if and only if $|\{i \in N : x R_i y\}| \geq |\{i \in N : y R_i x\}|$. This rule clearly satisfies weak Pareto, independence of irrelevant alternatives, and non dictatorship, but violates monotonicity.

Rule 2. Let $d \in N$. For all $x, y \in X$, let $x R_0 y$ if and only if $x R_d y$. This rule clearly satisfies monotonicity, weak Pareto, and independence of irrelevant alternatives, but violates non-dictatorship.

Rule 3. Let \mathcal{R}^T be the set of preference relations such that $R \in \mathcal{R}^T$ and $R' \succ R$ implies that $R' \in \mathcal{R}^T$. If $R_1, R_2 \in \mathcal{R}^T$, let $f(R_1, \dots, R_n) = R_1$, otherwise, let $f(R_1, \dots, R_n) = R_2$. This rule satisfies monotonicity, weak Pareto, and non dictatorship, but violates independence of irrelevant alternatives.

Rule 4. For all $x, y \in X$, let $x R_0 y$. This rule clearly satisfies monotonicity, independence of irrelevant alternatives, and nondictatorship, but violates weak Pareto. \square

References

- ARROW, K. J. (1963): *Social Choice and Individual Values*, Yale University Press, second ed.
- BARTHELEMY, J.-P. (1982): “Arrow’s theorem: Unusual domains and extended codomains,” *Mathematical Social Sciences*, 79–89.
- BATRA, R. AND K. PATTANAİK (1972): “Transitive multi-stage majority decisions with quasi-transitive individual preferences,” *Econometrica*, 40, 1121–1135.
- BAUCELLS, M. AND L. S. SHAPLEY (2008): “Multiperson utility,” *Games and Economic Behavior*, 62, 329–347.
- BLACK, D. (1948a): “The Decisions of a Committee Using a Special Majority,” *Econometrica*, 16, 245–261.
- (1948b): “On the Rationale of Group Decision-Making,” *Journal of Political Economy*, 56, 23–34.

- BLAIR, D. H. AND R. A. POLLAK (1982): “Acyclic Collective Choice Rules,” *Econometrica*, 50, 931–943.
- BOSSERT, W. AND K. SUZUMURA (2007): “Domain Closedness Conditions and Rational Choice,” *Order*, 24, 75–88.
- CATO, S. (2011): “Social choice without the Pareto principle: a comprehensive analysis,” *Soc Choice Welf*.
- CHEREPANOV, V., T. FEDDERSEN, AND A. SANDRONI (2008): “Rationalization,” Mimeo, Northwestern University.
- DASGUPTA, P. AND E. MASKIN (2008): “On the Robustness of Majority Rule,” *Journal of the European Economic Association*, 6, 949–973.
- DE CLIPPEL, G. (2012): “Behavioral Implementation,” Working paper.
- GIBBARD, A. (1969): “Social Choice and the Arrow Conditions,” Unpublished manuscript.
- INADA, K. (1969): “The simple majority decision rule,” *Econometrica*, 37, 490–506.
- MANZINI, P. AND M. MARIOTTI (2007): “Sequentially rationalizable choice,” *American Economic Review*, 97, 1824–1839.
- MAS-COLELL, A. AND H. SONNENSCHNEIN (1972): “General Possibility Theorems for Group Decisions,” *Review of Economic Studies*, 39, 185–192.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. OZBAY (forthcoming): “Revealed attention,” *American Economic Review*.
- MONJARDET, B. (1978): “An Axiomatic Theory of Tournament Aggregation,” *Mathematics of Operations Research*, 3, 334–351.
- PINI, M. S., F. ROSSI, K. B. VENABLE, AND T. WALSH (2009): “Aggregating Partially Ordered Preferences,” *Journal of Logic and Computation*, 19, 475–502.
- RUBINSTEIN, A. AND Y. SALANT (forthcoming): “Eliciting welfare preferences from behavioural data sets,” *Review of Economic Studies*.
- SAVAGE, L. J. (1972): *The Foundation of Statistics*, Dover Books on Mathematics Series, Dover Publications.
- SCHICK, F. (1969): “Arrow’s Proof and the Logic of Preference,” *Philosophy of Science*, 36, 127–144.
- SEN, A. (1969): “Quasi-Transitivity, Rational Choice and Collective Decisions,” *Review of Economic Studies*, 36, 381–393.

- (1977): “Social Choice Theory: A Re-Examination,” *Econometrica*, 45, 53–88.
- SEN, A. K. (1964): “Preferences, Votes and the Transitivity of Majority Decisions,” *Review of Economic Studies*, 31, 163–165.
- (1966): “A possibility theorem on majority decisions,” *Econometrica*, 34, 491–499.
- (1970): *Collective Choice and Social Welfare*, Holden-Day, Inc.
- SUZUMURA, K. (1983): *Rational choice, collective decisions, and social welfare*, Cambridge University Press.
- TVERSKY, A. (1969): “Intransitivity of preferences,” *Psychological Review*, 76, 31–48.