DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

AN AXIOMATIC THEORY OF POLITICAL REPRESENTATION

Christopher P. Chambers

SOCIAL SCIENCE WORKING PAPER 1218

March 2005

An axiomatic theory of political representation

Christopher P. Chambers

Abstract

We discuss the theory of voting rules which are immune to gerrymandering. Our approach is axiomatic. We show that any rule that is unanimous, anonymous, and representative consistent must decide a social alternative as a function of the proportions of agents voting for each alternative, and must either be independent of this proportion, or be in one-to-one correspondence with the proportions. In an extended model in which voters can vote over elements of the unit interval, we introduce and characterize the quasi-proportional rules based on unanimity, anonymity, representative consistency, strict monotonicity, and continuity. We show that we can always (pointwise) approximate a single-member district quota rule with a quasi-proportional rule. We also establish that upon weakening strict monotonicity, the generalized target rules emerge.

JEL classification numbers: D63, D70

Key words: gerrymandering, representative systems, proportional representation, social choice, quasi-arithmetic means

An axiomatic theory of political representation *

Christopher P. Chambers

1 Introduction

In representative democracies, there are two well-accepted methods of assigning representatives to districts. One, the single-member, or winner-take-all method, assigns a unique representative to each district of agents, as some function of the agents' votes. The other commonly used method is the method of proportional representation. In this method, a unique representative is not assigned to each district; rather, a collection of representatives is assigned to each district in proportion (approximately) equal to the number of votes each of the parties received in the districts.

The two ideals each have their benefits. The winner-take-all method has the benefit of giving each district a unique representative that respects its "collective interest." The proportional method accurately reflects the composition of votes received for different alternatives. In particular, under proportional representation systems, the benefit to strategically constructing districts in order to influence the outcome of a vote is significantly reduced.

We study an abstract theory of representation. Our particular interest is in rules and systems of representation for which there is no benefit to strategically constructing districts. We know of at least one system (proportional representation). Our goal is to obtain a broader understanding of these systems. We ask the following question: "For which voting rules and which systems of representation is it without loss of generality to group agents into voting districts?"

To this end, our formal model features a set of alternatives, over which voters vote. We imagine that only the votes that voters submit for a particular alternative are observed. The underlying preferences and strategic behavior that leads to these votes is outside of the formal model. We focus on the study of "rules." A rule recommends a unique alternative for every possible set of voters and every possible list of votes that the voters may submit. To address the question posed in the previous paragraph, we need to formalize what we mean by requiring voters to vote in districts.

*I would like to thank participants of the 2004 Meeting of the Society of Economic Design, and from the joint Harvard/MIT Economic Theory seminar

Suppose we have given a rule, and a list of votes that voters have submitted. Suppose that the voters have been grouped into districts. One can apply the rule to each district separately (this makes sense, as a rule is defined on every possible set of voters). The alternative selected by the voting rule for the district and the profile of votes is said to be the representative alternative of the district. This means that each voter in that district can be viewed as if she had voted for the representative alternative. By replacing each voters' vote by their representative alternative, we obtain a new vote profile. These "representative votes" can then be aggregated to determine the winning alternative for society. This is an example of "indirect voting." Similar notions have been studied by Murakami [13, 14], Fishburn [7, 8], and Fine [5].

For a typical rule, the way that the voters are partitioned into districts influences the outcome of the vote. When this can happen, we say that "gerrymandering" is possible. The property of being immune to gerrymandering is hereafter referred to as "representative consistency" of a rule.

As we envision the representative alternative for a district as taking all of the votes for that district, the notion of representative consistency implicitly refers to a winner-take-all system. We explain later how this framework also includes proportional representation as a special case.

In another work [3], we show that when society faces a decision among a finite set of alternatives, any rule which is democratic (in the sense of being anonymous and reflecting the will of the people when a unanimity of voters vote for a certain alternative) and representative consistent exhibits pathologies. Thus, such a rule must be a type of "unanimity rule," whereby alternatives are partially ordered, and any agent may veto an alternative with an alternative that is ranked more highly.

Our first main result is a related statement dealing with decisions that may feature an infinite number of alternatives. It states the following. Fix any pair of alternatives, and consider those environments in which these are the only two alternatives to receive votes. The result states that there are exactly two possibilities for such environments. One possibility is that the the alternative selected for such environments is independent of the vote profile (thus, it is constant as a function of the vote profile). The other possibility is that the rule is a function only of the proportion of voters voting for each alternative, and that there exists a one-to-one correspondence between this proportion and the set of alternatives.

One immediate implication of this result is that the set of alternatives which can be generated when voters vote for these two alternatives must either be a singleton or must be infinite, so that infinite sets of alternatives must be permitted if we are to rule out pathologies. Thus, *the only* possible method of avoiding this type of pathological rule is the introduction of new alternatives into the formal model.

Proportional representation relies on just such an enlargement of the set of alternatives. We know that systems of proportional representation, at least when perfectly implemented, are well-defined and independent of the way that districts are drawn (for an interesting axiomatic study of the integer problem in proportional representation, see Balinski and Young [2]). Our main insight is that, in systems of proportional representation, the outcome of a vote is not a single alternative, but is instead a *composition of a governing body.*

To explore this idea, we study an environment with an underlying binary social decision. We extend the traditional social choice model to allow voters to vote for objects such as: "a governing body which is composed 57% of agents in favor of alternative 0, and 43% of agents in favor of 1." We do not envision an environment in which voters actually vote for such constructs. But these extended alternatives will be important in a later stage, when *representative alternatives at the district level* are aggregated. Let us here emphasize that the extended set of alternatives need not be interpreted as compositions of a governing body. Indeed, this is one benefit of an abstract approach. Another natural interpretation is that the extended alternatives are lotteries over the two degenerate alternatives. Moreover, one may even choose to think of the extended set of alternatives as forming a classical unidimensional Euclidean policy space. All of these interpretations are discussed in our analysis.

Are there rules that are both immune to gerrymandering and democratic over the extended set of alternatives? A broad class of such rules is well-known from the mathematics literature. Define the "proportional rule" as that rule which simply takes the arithmetic mean of the votes that voters submit. The proportional rule corresponds to a system of proportional representation. However, there is no reason to think the arithmetic mean is special. In fact, any "quasi-arithmetic mean" will work just as well. Define a *quasi-proportional rule* as a rule that takes a quasi-arithmetic mean over all votes received.

The abstraction away from standard proportional representation may appear at first to be a mathematical exercise. However; there are important practical implications of the analysis. Recall that there are essentially no voting rules which are representative consistent with single-member district systems. To this end, we study "how close" we can come to a single-member district system, and still use a rule which is immune to gerrymandering.

It turns out that we can come very close. Within the class of quasi-proportional rules, we can construct a *sequence* of voting rules which are immune to gerrymandering, but which "converge" to majority rule with single-member districts, in a formal sense. The benefit of such a result from an institutional design standpoint is that one can construct a system which is completely immune to gerrymandering, but for which each district "almost" gets its own representative. In fact, we show that this approximation result actually holds for any "quota rule."[1] This is another of our primary results; that by extending the set of alternatives, one may construct voting rules that are immune to

[1]A quota rule is a rule for which there exists some status quo alternative, and a **quota** $q \in [0, 1]$, such that alternative 1 wins if the proportion of agents voting for 1 is greater than or equal to q.

gerrymandering but that are as close to majority rule with single-member districts as we like. Here, it is most reasonable to imagine that the "extended alternatives" over which agents vote are lotteries over the pair of degenerate alternatives. The result on the approximation of majority and quota rules by voting rules which are immune to gerrymandering is perhaps our simplest result, but arguably the most important.

Unfortunately, these types of results do not hold as we move away from binary decisions to decisions involving more than two alternatives. In fact, we show that essentially the only natural method of representation in this more general environment is the proportional rule.

We extend our analysis even further. Imagine a single-dimensional spatial model, in which agents may have preferences that are single-peaked. We discuss a family of rules that are non-pathological and immune to gerrymandering. A characteristic of the quasi-proportional rules is that they are "strictly monotonic," which has the implication that any voter can influence the outcome of the vote simply by changing her vote. There are many reasons that we may not desire a rule to be strictly monotonic–one natural reason is that it admits the possibility of strategic behavior of voters. That said, we weaken the idea of strict monotonicity to monotonicity–this simply states that a rule behaves non-perversely, so that if all agents vote for more representation for alternative 1, alternative 1 should get weakly more representation. A characterization of a broad family of democratic, representative consistent, monotonic, and continuous rules (also known from the mathematics literature, see Fodor and Marichal [9]) is provided. These rules simultaneously generalize the quasi-proportional rules, as well as the "target rules," introduced by Thomson [16] and Ching and Thomson [4].

Section 2 introduces the formal model. Section 3 provides a fundamental result motivating the remaining part of the study. Section 4 discusses our model of proportional representation. Section 5 concludes.

2 The model

Let X be an arbitrary set of alternatives. There is an infinite set of potential agents, which we without loss of generality index by the natural numbers $\mathbb{N}$. At any given time, we will only consider finite subsets of $\mathbb{N}$. The set of finite subsets of $\mathbb{N}$ is denoted by $\mathcal{N}$. A rule is a function $r : \bigcup_{N\in\mathcal{N}} X^N \to X$, recommending for each society N of voters, and each vote profile $x \in X^N$ some alternative for society.

The following conditions were studied in Chambers [3].

2.1 Democratic principles

The first property we discuss in this section states that a rule should respect the "will of the people" when this "will" is unambiguous. It is an extremely weak axiom.

For all $N \in \mathcal{N}$ and all $x \in X$, let x^N be a vector in X^N such that for all $i \in N$, $x_i^N = x$. For all $N \in \mathcal{N}$, all $x \in X^N$, and all $M \subset N$, let x_M be the restriction of x to X^M.

Unanimity: For all $N \in \mathcal{N}$ and all $x \in X$, $f\left(x^N\right) = x$.

The next axiom states that a rule should be ignorant of the names of agents.

Anonymity: For all $N, N' \in \mathcal{N}$ such that $|N| = |N'|$, all bijections $\sigma : N \to N'$, and all $x \in X^N$ and $x' \in X^{N'}$ such that for all $i \in N$, $x_i = x'_{\sigma(i)}$, $f(x) = f(x')$.

2.2 Representative consistency and gerrymandering

Informally, representative consistency states that for any population of agents and any collection of votes, it is without loss of generality to partition the set of agents into districts, find the choice for each district, and then treat each district as if each agent in the district had voted for the outcome selected for the district.

Representative consistency: For all $N \in \mathcal{N}$, all partitions $\{N_1, ..., N_K\}$ of N, and all $x \in X^N$, $f(x) = f\left(f(x_{N_1})^{N_1}, ..., f(x_{N_K})^{N_K}\right)$.

Under the unanimity principle, representative consistency is equivalent to the stricter statement that for all $N \in \mathcal{N}$, all $M \subset N$, and all $x \in X^N$, $f(x) = f\left(f(x_M)^M, x_{N \setminus M}\right)$. This latter version of representative consistency is more useful in the proofs of theorems.

Any anonymous rule can be specified without reference to the specific names of agents. In the proofs of results in which anonymity plays a role, we often exploit this fact without mention, disregarding the variable N.

The following axiom can be interpreted as meaning that only the *proportions* of votes received for each alternative are used in determining the social alternative.

Let m be an integer, let $N \in \mathcal{N}$, and let $x \in X^N$. Let $N' \in \mathcal{N}$ be such that $|N'| = m|N|$. A vector $x' \in X^{N'}$ is an **m-replica of x** if there exists a partition of N' into m sets of size $|N|$, say $\{N_1, ..., N_m\}$ such that for all N_i, there exists a bijection $\sigma_i : N \to N_i$ so that for all $j \in N$, $x_j = x'_{\sigma(j)}$. For all $N \in \mathcal{N}$, $x \in X^N$, $m \in \mathbb{N}$, x^m denotes an m-replica of x.

Replication invariance: Let m be an integer. Let $N \in \mathcal{N}$ and let $x \in X^N$. Let x' be an m-replica of x. Then $f(x') = f(x)$.

3 A theorem of the alternative for representation

The following trivial observation is useful:

Lemma 1: If a rule satisfies unanimity, anonymity, and representative consistency, then it satisfies replication invariance.

Proof: Let $N \in \mathcal{N}$ and let $x \in X^N$. Let x' be an m-replica of x. Then by definition of x', $f(x') = f\left(\underbrace{x, ..., x}_{m}\right)$. By representative consistency, $f(x') = f\left(\underbrace{f(x), ..., f(x)}_{m|N|}\right)$. By unanimity, $f\left(\underbrace{f(x), ..., f(x)}_{m|N|}\right) = f(x)$. Thus $f(x') = f(x)$.■

The following theorem gives us a general result on the structure of democratic rules which are representative consistent. Let $x, y \in X$. For $N \in \mathcal{N}$, say that $\mathbf{z} \in \{\mathbf{x}, \mathbf{y}\}^N$ **is an $\{\mathbf{x}, \mathbf{y}\}$-profile** if there exists $i \in N$ such that $z_i = x$ and $j \in N$ such that $z_j = y$. Thus, z is an $\{x, y\}$-profile if all voters vote for either x or y, and at least one voter votes for x and one votes for y.

The theorem is a "Theorem of the alternative." It states that, in an abstract environment, if a rule satisfies our primary axioms, then for all pairs $x, y \in X$ there are two (mutually exclusive) possibilities. The first, pathological, possibility is that the rule is constant on the set of all $\{x, y\}$-profiles. Such a rule does not recognize the proportions of agents voting for each alternative x and y. Such rules are investigated in Chambers [3]. The second possibility is that the rule is a one-to-one correspondence between the set of proportions of voters voting for x, and the set of alternatives. This second possibility requires that the set of alternatives be infinite (as the set of proportions is clearly infinite); but it also requires that a rule can only be based on the set of proportions of voters voting for one alternative over another. Moreover, it requires that a rule completely discriminate among proportions. We first present the proof, then an informal description.

Theorem 1 (Theorem of the Alternative): Suppose that a rule f satisfies unanimity, anonymity, and representative consistency. Then for all pairs $x, y \in X$, one and only one of the following is true. i) f is constant on the set of $\{x, y\}$-profiles, ii) there exists a one-to-one function $g^{(x,y)} : (0, 1) \cap \mathbb{Q} \rightarrow X$ such that for all $N \in \mathcal{N}$, if $z \in \{x, y\}^N$ is an $\{x, y\}$-profile, $f(z) = g^{(x,y)}\left(\frac{|\{i \in N : z_i = x\}|}{|N|}\right)$.

Proof: Step 1: Construction of an auxiliary function

First, we construct an auxiliary function $h : \mathbb{Q}^2_{++} \to X$ in the following manner. For all pairs $(\alpha, \beta) \in \mathbb{Q}^2_{++}$, we may write $\alpha = \frac{m(\alpha)}{n}$, and $\beta = \frac{m(\beta)}{n}$, where $m(\alpha), m(\beta)$ are natural numbers greater than zero, and n is a natural number greater than zero. We define $h(\alpha, \beta) = f\left(x^{m(\alpha)}, y^{m(\beta)}\right)$. Although the representation of α and β in terms of ratios of natural numbers is not unique, it is unique up to scalar multiplication. Therefore, by replication invariance, h is well-defined.

Step 2: Establishing "additivity" of the auxiliary function

We establish that for all $(\alpha, \beta), (\alpha', \beta') \in \mathbb{Q}^2_{++}$, if $h(\alpha, \beta) = h(\alpha', \beta')$, then $h(\alpha + \alpha', \beta + \beta') = h(\alpha, \beta)$. To this end, suppose that (α, β) and (α', β') are such that $h(\alpha, \beta) = h(\alpha', \beta')$. Label $z \equiv h(\alpha, \beta)$. There exists some n large enough so that $\alpha = \frac{m(\alpha)}{n}, \beta = \frac{m(\beta)}{n}, \alpha' = \frac{m(\alpha')}{n}, \beta' = \frac{m(\beta')}{n}$. Thus, $\alpha + \alpha' = \frac{m(\alpha)+m(\alpha')}{n}$ and $\beta + \beta' = \frac{m(\beta)+m(\beta')}{n}$. By definition of h,

$$h(\alpha + \alpha', \beta + \beta') = f\left(x^{m(\alpha)+m(\alpha')}, y^{m(\beta)+m(\beta')}\right).$$

Rewriting,

$$f\left(x^{m(\alpha)+m(\alpha')}, y^{m(\beta)+m(\beta')}\right) = f\left(\left(x^{m(\alpha)}, y^{m(\beta)}\right), \left(x^{m(\alpha')}, y^{m(\beta')}\right)\right).$$

By representative consistency,

$$\begin{aligned} & f\left(\left(x^{m(\alpha)}, y^{m(\beta)}\right), \left(x^{m(\alpha')}, y^{m(\beta')}\right)\right) \\ = \; & f\left(f\left(x^{m(\alpha)}, y^{m(\beta)}\right)^{m(\alpha)+m(\beta)}, f\left(x^{m(\alpha')}, y^{m(\beta')}\right)^{m(\alpha')+m(\beta')}\right). \end{aligned}$$

But the preceding is $f\left(z^{m(\alpha)+m(\beta)}, z^{m(\alpha')+m(\beta')}\right)$, so that by unanimity, $f\left(z^{m(\alpha)+m(\beta)}, z^{m(\alpha')+m(\beta')}\right) = z = h(\alpha, \beta)$.

The preceding paragraph tells us that the equivalence classes for the function h are $\mathbb{Q}$-convex cones, closed under addition and rational scalar multiplication. This last property will be called **rational homogeneity** of h.

Step 3: Establishing "translation invariance" of the auxiliary function

Next, we claim that for all $(\alpha, \beta), (\alpha', \beta') \in \mathbb{Q}^2_{++}$, if $\alpha + \beta = \alpha' + \beta'$, and $h(\alpha, \beta) = h(\alpha', \beta')$, then for all $(\alpha'', \beta'') \in \mathbb{Q}^2_{+}$ (here, pairs of nonnegative rational numbers), $h(\alpha + \alpha'', \beta + \beta'') = h(\alpha' + \alpha'', \beta' + \beta'')$. We will call this property **translation invariance of equivalence classes**. To see this, again note that we may choose

n large so that $\alpha = \frac{m(\alpha)}{n}$, $\beta = \frac{m(\beta)}{n}$, $\alpha' = \frac{m(\alpha')}{n}$, $\beta' = \frac{m(\beta')}{n}$, $\alpha'' = \frac{m(\alpha'')}{n}$, and $\beta'' = \frac{m(\beta'')}{n}$. Clearly, $m(\alpha) + m(\beta) = m(\alpha') + m(\beta')$. Write

$$h(\alpha + \alpha'', \beta + \beta'') = f\left(x^{m(\alpha)+m(\alpha'')}, y^{m(\beta)+m(\beta'')}\right).$$

Rewriting, we obtain

$$f\left(x^{m(\alpha)+m(\alpha'')}, y^{m(\beta)+m(\beta'')}\right) = f\left(\left(x^{m(\alpha)}, y^{m(\beta)}\right), \left(x^{m(\alpha'')}, y^{m(\beta'')}\right)\right).$$

By applying representative consistency, we obtain

$$\begin{aligned} & f\left(\left(x^{m(\alpha)}, y^{m(\beta)}\right), \left(x^{m(\alpha'')}, y^{m(\beta'')}\right)\right) \\ = \ & f\left(f\left(x^{m(\alpha)}, y^{m(\beta)}\right)^{m(\alpha)+m(\beta)}, f\left(x^{m(\alpha'')}, y^{m(\beta'')}\right)^{m(\alpha'')+m(\beta'')}\right). \end{aligned}$$

As

$$f\left(x^{m(\alpha)}, y^{m(\beta)}\right) = h(\alpha, \beta) = h(\alpha', \beta') = f\left(x^{m(\alpha')}, y^{m(\beta')}\right),$$

and as $m(\alpha) + m(\beta) = m(\alpha') + m(\beta')$, conclude

$$\begin{aligned} & f\left(f\left(x^{m(\alpha)}, y^{m(\beta)}\right)^{m(\alpha)+m(\beta)}, f\left(x^{m(\alpha'')}, y^{m(\beta'')}\right)^{m(\alpha'')+m(\beta'')}\right) \\ = \ & f\left(f\left(x^{m(\alpha')}, y^{m(\beta')}\right)^{m(\alpha')+m(\beta')}, f\left(x^{m(\alpha'')}, y^{m(\beta'')}\right)^{m(\alpha'')+m(\beta'')}\right). \end{aligned}$$

By representative consistency,

$$\begin{aligned} & f\left(f\left(x^{m(\alpha')}, y^{m(\beta')}\right)^{m(\alpha')+m(\beta')}, f\left(x^{m(\alpha'')}, y^{m(\beta'')}\right)^{m(\alpha'')+m(\beta'')}\right) \\ = \ & f\left(\left(x^{m(\alpha')}, y^{m(\beta')}\right), \left(x^{m(\alpha'')}, y^{m(\beta'')}\right)\right). \end{aligned}$$

Rewriting obtains

$$f\left(\left(x^{m(\alpha')}, y^{m(\beta')}\right), \left(x^{m(\alpha'')}, y^{m(\beta'')}\right)\right) = f\left(x^{m(\alpha')+m(\alpha'')}, y^{m(\beta')+m(\beta'')}\right).$$

This is in turn equal to $h(\alpha' + \alpha'', \beta' + \beta'')$. Hence $h(\alpha + \alpha'', \beta + \beta'') = h(\alpha' + \alpha'', \beta' + \beta'')$.

Step 4: Using the auxiliary function to establish the theorem

Next, to show that either i) or ii) in the statement of the Theorem must be true, suppose that they are both false. They are clearly mutually exclusive statements. As ii) is false, there exist $\alpha, \beta \in \mathbb{Q} \cap (0,1)$ such that $\alpha < \beta$, and $h(\alpha, 1-\alpha) = h(\beta, 1-\beta)$. By the preceding arguments, for all $\gamma \in \mathbb{Q} \cap (\alpha, \beta)$, $h(\gamma, 1-\gamma) = h(\alpha, 1-\alpha)$. As i) is false, h is nonconstant, so there exists some $\gamma < 1$ (without loss of generality $\gamma > \beta$) such that $h(\gamma, 1-\gamma) \neq h(\beta, 1-\beta)$.

Let $\gamma^* \equiv \inf\{\gamma > \beta : h(\gamma, 1-\gamma) \neq h(\beta, 1-\beta)\}$ and let $x^* = \frac{\gamma^*-\beta}{1-\gamma^*}$. Clearly, $x^* \geq 0$. In particular, $\frac{\beta+x^*}{1+x^*} = \gamma^*$. Moreover, $\frac{\alpha+x^*}{1+x^*} < \frac{\beta+x^*}{1+x^*}$. By continuity of $\frac{\alpha+x^*+y}{1+x^*+y}$ in y, there exists $x > 0$ such that $\frac{\alpha+x^*+x}{1+x^*+x} \in \mathbb{Q}$ and for which $\frac{\alpha+x^*+x}{1+x^*+x} < \frac{\beta+x^*}{1+x^*} = \gamma^*$. In particular, $\frac{\alpha+x^*+x}{1+x^*+x} < \gamma^* - \varepsilon$ for some $\varepsilon > 0$. Further, $\frac{\alpha+x^*+x}{1+x^*+x} > \alpha$, so that by definition of x^* and $\mathbb{Q}$-convexity of h, $h\left(\frac{\alpha+x^*+x}{1+x^*+x}, \frac{1-\alpha}{1+x^*+x}\right) = h(\beta, 1-\beta)$.

Next, as $\frac{1-\alpha}{1+x^*+x} = 1 - \frac{\alpha+x^*+x}{1+x^*+x}$, $\frac{1-\alpha}{1+x^*+x} \in \mathbb{Q}$ and hence as $1-\alpha \in \mathbb{Q}$, we conclude that $1 + x^* + x \in \mathbb{Q}$. Therefore, by rational homogeneity, $h\left(\frac{\alpha+x^*+x}{1+x^*+x}, \frac{1-\alpha}{1+x^*+x}\right) = h(\alpha + x^* + x, 1-\alpha)$. As $h(\alpha, 1-\alpha) = h(\beta, 1-\beta)$, translation invariance of equivalence classes implies that $h(\alpha + x^* + x, 1-\alpha) = h(\beta + x^* + x, 1-\beta)$. Lastly, by rational homogeneity, $h(\beta + x^* + x, 1-\beta) = h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right)$. Putting together the equalities, we conclude $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) = h(\beta, 1-\beta)$.

However, since $\frac{\beta+x^*}{1+x^*} < 1$, we can conclude that $\frac{\beta+x^*+x}{1+x^*+x} > \frac{\beta+x^*}{1+x^*} = \gamma^*$. We conclude that $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) \neq h(\beta, 1-\beta)$. To see this, suppose, by means of contradiction, that $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) = h(\beta, 1-\beta)$. Then, by $\mathbb{Q}$-convexity of h, for all $\gamma \in \left(\beta, \frac{\beta+x^*+x}{1+x^*+x}\right)$, $h(\gamma, 1-\gamma) = h(\beta, 1-\beta)$. But $\gamma^* \in \left(\beta, \frac{\beta+x^*+x}{1+x^*+x}\right)$. It follows by definition of γ^* that there exists some $\gamma \in \left(\beta, \frac{\beta+x^*+x}{1+x^*+x}\right)$ for which $h(\gamma, 1-\gamma) \neq h(\beta, 1-\beta)$, a contradiction. Hence, $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) \neq h(\beta, 1-\beta)$.

We therefore conclude both $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) = h(\beta, 1-\beta)$ and $h\left(\frac{\beta+x^*+x}{1+x^*+x}, \frac{1-\beta}{1+x^*+x}\right) \neq h(\beta, 1-\beta)$. This is a contradiction. ■

While the proof of Theorem 1 appears complicated, in fact the idea is very simple. The function h as constructed in the proof takes as arguments pairs of positive rational numbers (p, q). The output of h for any such pair is an alternative; it is the unique alternative recommended for a society for which a fraction $\frac{p}{p+q}$ of the agents vote for x and a fraction $\frac{q}{p+q}$ vote for y. It is easily verified that the equivalence classes for this function are "convex" cones. The claim in Theorem 1 is that either there is only one equivalence class of h, or that there exist an infinite number of equivalence classes, and each equivalence class is simply a ray. When we assume that the statement of Theorem 1 is false, we are assuming that there exists an equivalence class of h which is neither a ray and which is also not the entire domain of h. Figure 1 depicts such a cone, and it depicts two points, $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ lying in the same cone. A key step in the proof of Theorem 1 is the verification that if $h(\alpha, 1-\alpha) = h(\beta, 1-\beta)$, then adding the same vector to both $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ does not change this equality. However, Figure 1 clearly demonstrates that for an appropriate choice of $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$, this property is violated. Here, we have found a vector, which when added to $(\alpha, 1-\alpha)$, results in a vector which remains inside the cone, and when added to $(\beta, 1-\beta)$ results in a vector which lies outside of the cone. Thus, the addition of the same vector to both

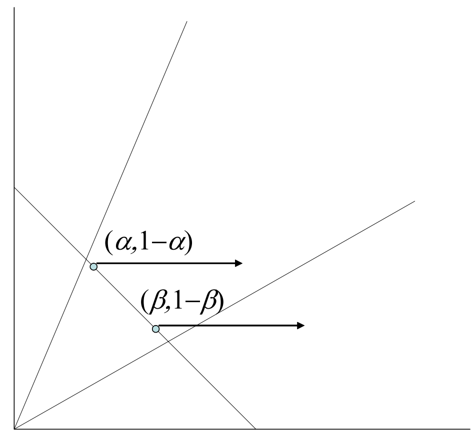

Figure 1: The proof of Theorem 1

$(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ results in two vectors lying in different equivalence classes of h; contradicting the property mentioned above.

The preceding theorem is illustrated by the following natural example.

Example 1: Let $X = \mathbb{R}_+$. For all $N \in \mathcal{N}$ and for all $x \in X^N$, define $f(x) = \sqrt[|N|]{\prod_{i \in N} x_i}$. Thus, f is the geometric mean of the votes received. One can easily verify that f satisfies unanimity, anonymity, and representative consistency. Moreover, f is not pathological. Let $x, y \in (0, +\infty)$. Let $g^{(x,y)} : \mathbb{Q} \cap (0,1) \to \mathbb{R}$ as in the preceding proposition as follows: $g(\alpha) = x^\alpha y^{(1-\alpha)}$. For all $\{x,y\}$-profiles z, $f(z) = \sqrt[|N|]{x^{|\{i \in N : z_i = x\}|} y^{|\{i \in N : z_i = y\}|}} = x^{\frac{|\{i \in N : z_i = x\}|}{|N|}} y^{1 - \frac{|\{i \in N : z_i = x\}|}{|N|}} = g^{(x,y)}\left(\frac{|\{i \in N : z_i = x\}|}{|N|}\right)$. For $\{x,y\}$-profiles, then, f satisfies condition ii) of the proposition. However, if either $x = 0$ or $y = 0$, then for all $\{x,y\}$-profiles z, $f(z) = 0$. Thus, if either $x = 0$ or $y = 0$, f satisfies condition i) of the proposition.

The proposition also tells us that if we only have a finite number of alternatives available that we would like to have as the social choice for an $\{x,y\}$-profile, then f must be constant on all $\{x,y\}$-profiles. The immediate implication of this result is that a society voting over a finite number of alternatives that wants to eliminate the possibility of gerrymandering *must introduce an infinitude of new alternatives over which to vote*. At first, this seems like a ridiculous idea; but we argue that this is exactly what takes place in the real world. In systems of proportional representation, the outcome of a vote is *not* an alternative per se, but a *composition of a governing body*. There are an infinite number of possible compositions of governing bodies. While voters do not actually get to vote for any composition that they like, allowing them to do so would not change the fact that gerrymandering is impossible for systems of proportional representation.

Indeed; there are other ways of introducing alternatives into the model. One can envision taking lotteries over the two alternatives in question; thus producing a continuum. Other types of mixing are possible. For example, if society is voting over how much, out of two levels, to spend on a public project, it is simple to allow the society to spend any amount in between.

4 Infinite sets of alternatives

4.1 The quasi-arithmetic means and quasi-proportional representation

Building on the results in the preceding section, we use this section to explore the implications of allowing the set of alternatives to be infinite. One of our primary aims is to discuss a notion of proportional representation.

In arbitrary infinite sets, many bizarre rules can be constructed which satisfy the axioms. A general characterization of the family does not seem possible at this time. However; we will be content in this section to study the very special case of the unit interval (and later on any finite-dimensional simplex).

We consider a binary social choice model; there are two alternatives or positions that society must decide between, say $\{0,1\}$. The novelty here is that agents are not restricted to vote for elements of $\{0,1\}$; they may also vote for any element $x \in [0,1]$. The element x here can be interpreted as the desired *proportion* of agents in some governing body that support alternative 1 (naturally, $1-x$ support alternative 0).

Of course, we should not expect voters to vote for elements in the interior of $(0,1)$. However, recall the two-stage aggregation procedure discussed above. It is natural to allow the outcome of a vote in a given district to be a composition of agents in a governing body; this is exactly what proportional representation allows. As the same rule is used at various stages of aggregation, we allow a rule to take as input such objects.

Our goal is to understand which forms of representation cannot be gerrymandered (among democratic voting rules). Of course, it is well-known that proportional representation is immune to gerrymandering. Here, in our model, proportional representation is defined as the following rule: for all $N \in \mathcal{N}$ and all $x \in [0,1]^N$, $f(x) = \sum_{i \in N} \frac{x_i}{|N|}$. Thus, f is simply the arithmetic mean of those alternatives which receive votes.

A moment's thought establishes that there is nothing special about the arithmetic mean. Other notions of mean have been defined in the mathematics literature (called

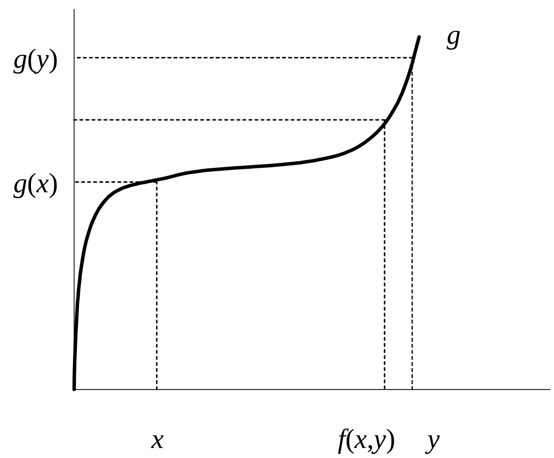

Figure 2: A quasi-proportional rule

the quasi-arithmetic means). Thus, let $g : [0,1] \to \mathbb{R}$ be a continuous, strictly increasing function. The quasi-arithmetic mean (with respect to g) is defined as follows: for all $N \in \mathcal{N}$, for all $x \in [0,1]^N$, $f(x) = g^{-1}\left(\frac{\sum_{i \in N} g(x_i)}{|N|}\right)$. The quasi-arithmetic means satisfy all of the axioms we have posited. They can be interpreted as transforming proportion space into another space, taking the average, and then transforming back to the original proportion space. Such means define a society-specific notion of average. In this context, we will call a rule which is generated by a quasi-arithmetic mean a **quasi-proportional rule**.

Figure 2 depicts a typical quasi-proportional rule. There is a continuous and strictly increasing function g. Fix any two points $x, y \in [0,1]$. We compute $f(x,y)$ as follows. First, find each of $g(x)$ and $g(y)$. Then, take the average of these two points. Finally, $f(x,y)$ is found of the inverse (under g) of this average.

Still other rules exist which satisfy all of the axioms. For example, the "positional dictatorship," defined as $f(x) = \min_{i \in N} \{x_i\}$ for all $N \in \mathcal{N}$ and all $x \in [0,1]^N$ also satisfies all of our axioms. In fact, the positional rule can be understood as a natural generalization of a unanimity-type rule with 0 as a status quo.

The preceding rules all have several characteristics in common. Firstly, they are continuous in all parameters. Importantly, they are also "monotonic," in the sense that if all agents' votes move weakly to the right, then so does the recommended social alternative.

Continuity: For all $N \in \mathcal{N}$, the rule f is continuous over X^N.

Monotonicity: For all $N \in \mathcal{N}$, for all $x, y \in X^N$, if for all $i \in N$, $x_i \leq y_i$, $f(x) \leq f(y)$.

A stronger version of monotonicity is also useful.

Strict monotonicity: For all $N \in \mathcal{N}$, for all $x, y \in X^N$, if for all $i \in N$, $x_i \leq y_i$ and $x \neq y$, $f(x) < f(y)$.

The following well-known theorem characterizes all strictly monotonic and continuous rules satisfying the democratic and gerrymandering-proofness properties. Versions of the theorem were first proved by Kolmogorov [11], Nagumo [15], and de Finetti [6]. We will not give a proof of this well-known result; a standard reference is Aczél [1]. In a later section, we investigate the implications of weakening the strict monotonicity axiom.

Theorem 2: A rule satisfies unanimity, anonymity, representative consistency, continuity, and strict monotonicity if and only if it is a quasi-proportional rule.

How can a quasi-proportional rule actually be implemented? It depends on the interpretation of the set of alternatives, but if the set of alternatives is actually interpreted as the compositions of a governing body, then the most natural quasi-proportional rule to use is the proportional rule itself. Informally, to understand why, for each district of voters, the outcome of a vote is simply a composition of a governing body. Naturally, one would expect the representatives for this district to be in proportion to the recommended composition made by the rule for the district. In a governing body, another natural requirement is that the number of representatives in each district should be proportional to the populations from each district. These two requirements cannot both be met unless the proportional rule itself is used. Therefore, Theorem 2 establishes a foundation for using the proportional rule of representation when alternatives are interpreted as compositions of a governing body. However; as we mentioned in the introduction, other interpretations are certainly possible, and it is to another of these to which we now turn.

4.2 Using lotteries to approximate single-member district systems

It is well-known (by both academics and politicians) that majority rule with single-member districts is not representative consistent, and thus, gerrymandering is a commonplace phenomenon. What has not been discussed before, however, is the degree to which this phenomenon is robust. Here, we introduce a method by which a society can come as close as they like to a majority rule system with single-member districts, but for which gerrymandering has no effect. Let us again suppose that there are two degenerate alternatives, over which voters vote. Now; however, we extend the set of alternatives to include lotteries over the degenerate alternatives. While formally identical to the analysis conducted in the last section, the interpretation is very different here.

More formally, Theorem 2 can be used to establish a result on the approximation of single-member district systems by rules which are gerrymandering-proof. Suppose that we have reason to desire majority rule with single-member districts (and some anonymous tie-breaking rule). As is evidenced from Theorem 1, this rule is clearly susceptible to gerrymandering. However; we show that we can construct a quasi-proportional representative system which *approximates* the single-member district rule to an arbitrarily high degree, when the rule is restricted to the class of $\{0,1\}$-profiles. This reflects an environment in which all voters vote for degenerate alternatives (as in a general election, for example).

Let $N \in \mathcal{N}$ and let $x \in \{0,1\}^N$, so that x is a $\{0,1\}$-profile. For such a profile, majority rule is that rule for which $f^{maj}(x) = 1_{\left\{x: \frac{\sum_{i \in N} x_i}{|N|} \geq .5\right\}}$.[2] The claim is that there exists a sequence $\{f^m\}_{m=1}^{\infty}$ of quasi-proportional rules such that for all $N \in \mathcal{N}$ and all $x \in \{0,1\}^N$, $f^m(x) \to f^{maj}(x)$. This means that, for any profile of votes for which all voters vote only for degenerate alternatives, we can approximate the decision made by majority rule to an arbitrarily high degree.

This result is a result on pointwise approximation; thus, how close the rule is to majority rule is a function of the specific vote profile in question. However, more is true if we assume an ex-ante upper bound on the cardinality of the set of agents. In this scenario, we can fix a population size n and some degree of error ε, so that there exists a quasi-proportional rule such that for all societies with a population less than n who vote only for elements of $\{0,1\}$, the quasi-proportional rule recommends an alternative within ε of what majority rule with single-member districts would recommend. This follows immediately from the definition of pointwise convergence and the fact that the set of possible profiles (up to permutation) involving less than n voters is finite (when all voters vote for degenerate alternatives).

Formally, given $n > 0$ and $\varepsilon > 0$, there exists $M > 0$ so that for all $m \geq M$, all $N \in \mathcal{N}$ for which $|N| \leq n$, for all $x \in \{0,1\}^N$, $\|f^m(x) - f^{maj}(x)\| < \varepsilon$. Thus, the sequence $\{f^m\}_{m=1}^{\infty}$ converges uniformly on this restricted class of vote profiles. Given a maximal size of society, across all profiles for which all agents vote either for 0 or for 1, we can choose a gerrymandering-proof democratic voting rule that coincides with majority rule with single-member districts to an arbitrarily high degree. Our proof actually shows how to construct such a rule by providing an explicit analytical expression.

The theorem is actually not specific to majority rule, so we show how to prove it for the **quota rules**. A quota rule is a rule for which a fixed proportion of agents q is required in order for society to select the alternative 1; otherwise the alternative 0 is selected. To simplify the notation, we define the quota rules for all possible vote profiles. However, the theorem only applies for vote profiles where all voters vote for one of the two alternatives.

[2] The function 1 is the "indicator function," taking a value of 1 on the set and 0 otherwise. Note that this specification breaks ties in favor of alternative 1; this has no effect on the results.

The quota rules are parametrized by a value in $(0,1)$.[3] Thus, let $q \in (0,1)$. Define the quota rule f^q as follows. For all $N \in \mathcal{N}$, and for all $x \in [0,1]^N$,

$$f^q(x) \equiv \left\{ \begin{array}{c} 1 \text{ if } \frac{\sum_{i \in N} x_i}{|N|} \geq q \\ 0 \text{ if } \frac{\sum_{i \in N} x_i}{|N|} < q \end{array} \right\}.$$

Naturally, we could replace the weak inequality with strict and vice-versa.

Theorem 3 (Quota Rule Approximation): Let f^q be a quota rule. Then there exists a sequence $\{f^m\}_{m=1}^{\infty}$ of quasi-proportional rules such that for all $N \in \mathcal{N}$ and all $x \in \{0,1\}^N$, $f^m(x) \to f^q(x)$. Moreover, for all $n \in \mathbb{N}$ and $\varepsilon > 0$, there exists some quasi-proportional rule f such that for all $N \in \mathcal{N}$ for which $|N| \leq n$ and for all $x \in \{0,1\}^N$, $\|f(x) - f^q(x)\| < \varepsilon$.

Proof: Step 1: Establishing the pointwise convergence result

Let f^q be a quota rule. For all $m \in \mathbb{N}$, such that $1/m < q$ and $m > 2$, define the piecewise linear (in three pieces) function $h^m : [0,1] \to [0,1]$ by

$$h^m(x) \equiv \left\{ \begin{array}{c} (mq-1)\,x \text{ for } 0 \leq x < \frac{1}{m} \\ \frac{x}{m-2} + q - \left(\frac{m-1}{m}\right)\left(\frac{1}{m-2}\right) \text{ for } \frac{1}{m} \leq x \leq 1 - \frac{1}{m} \\ m(1-q)\,x + 1 - m(1-q) \text{ for } 1 - \frac{1}{m} < x \leq 1 \end{array} \right\}.$$

Each h^m is continuous and strictly monotonic. Let f^m be the quasi-proportional rule defined with the function h^m. We claim that the first statement in the claim of Theorem 3 holds with respect to the sequence $\{f^m\}_{m=1}^{\infty}$. To see this, let $N \in \mathcal{N}$ and $x \in \{0,1\}^N$, and suppose that $\frac{\sum_{i \in N} x_i}{|N|} \geq q$. In particular, then, as x consists solely of zeroes and ones, and as $h^m(0) = 0$ and $h^m(1) = 1$, we conclude that $\frac{\sum_{i \in N} h^m(x_i)}{|N|} \geq q$. Moreover, $(h^m)^{-1}(q) = 1 - \frac{1}{m}$, and as h^m is monotonic, so is its inverse; hence $(h^m)^{-1}\left(\frac{\sum_{i \in N} h^m(x_i)}{|N|}\right) \geq 1 - \frac{1}{m}$, so that $(h^m)^{-1}\left(\frac{\sum_{i \in N} h^m(x_i)}{|N|}\right) \to 1$.

Suppose next that $\frac{\sum_{i \in N} x_i}{|N|} < q$. In particular, then, as x consists solely of zeroes and ones, and as $h^m(0) = 0$ and $h^m(1) = 1$, we conclude that $\frac{\sum_{i \in N} h^m(x_i)}{|N|} < q - \eta$ for some $\eta > 0$ and all m. Thus, there exists an M large enough so that for all $m > M$, $\frac{\sum_{i \in N} h^m(x_i)}{|N|} < q - \frac{1}{m}$. But $(h^m)^{-1}\left(q - \frac{1}{m}\right) = \frac{1}{m}$. Hence, by monotonicity of $(h^m)^{-1}$, $(h^m)^{-1}\left(\frac{\sum_{i \in N} h^m(x_i)}{|N|}\right) \leq (h^m)^{-1}\left(q - \frac{1}{m}\right)$. Thus, $(h^m)^{-1}\left(\frac{\sum_{i \in N} h^m(x_i)}{|N|}\right) \to 0$.

Step 2: Establishing the uniform bound result

[3] We could actually allow $q = 0, 1$, but these correspond to unanimity rules. We know unanimity rules are already representative consistent. Moreover, the min and max rules for the extended model coincide with the unanimity rules. These rules are also representative consistent. Therefore, there is no need to discuss approximation in this case.

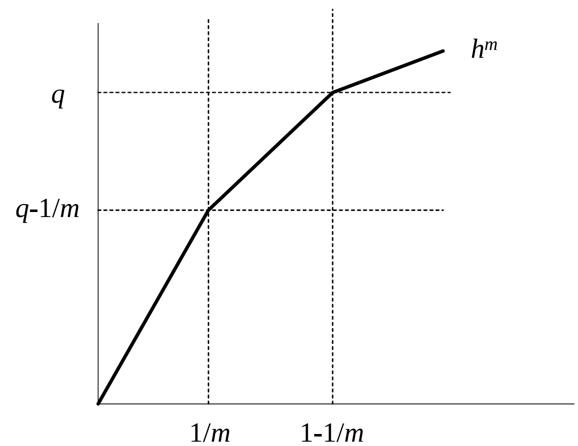

Figure 3: The proof of Theorem 3

To verify the second statement, let $n \in \mathbb{N} \cup \{0\}$ and let $\varepsilon > 0$. Let $\mathbb{Q}_n \equiv \left\{\frac{m}{k} : m, k \in \mathbb{N},\, k \leq n\right\}$. The set $\mathbb{Q}_n$ is finite, and we may order it $\mathbb{Q}_n = \{q_1, ..., q_L\}$, where $L < +\infty$. Let K satisfy $q_K < q \leq q_{K+1}$. Let M be any integer so that $\frac{1}{M} < \min\{\varepsilon, q - q_K\}$, and let $f = f^M$, as defined in the preceding part of the proof.

Let $N \in \mathcal{N}$ so that $|N| \leq n$, and let $x \in \{0,1\}^N$. If $\frac{\sum_{i \in N} x_i}{|N|} \geq q$, then in Step 1, $\frac{\sum_{i \in N} h^M(x_i)}{|N|} \geq q$, thus $\left(h^M\right)^{-1}\left(\frac{\sum_{i \in N} h^M(x_i)}{|N|}\right) \geq \left(h^M\right)^{-1}(q) = 1 - \frac{1}{M}$. Hence, $f(x) \geq 1 - \frac{1}{M}$ and $f^q(x) = 1$, so that $\|f(x) - f^q(x)\| \leq \frac{1}{M} < \varepsilon$.

Next, suppose that $\frac{\sum_{i \in N} x_i}{|N|} < q$. Then, as in Step 1, $\frac{\sum_{i \in N} h^M(x_i)}{|N|} < q$. In particular, $\frac{\sum_{i \in N} h^M(x_i)}{|N|} \in \mathbb{Q}_n$, so that $\frac{\sum_{i \in N} h^M(x_i)}{|N|} \leq q_K$. Therefore, $\frac{\sum_{i \in N} h^M(x_i)}{|N|} < q - \frac{1}{M}$. In particular, $\left(h^M\right)^{-1}\left(\frac{\sum_{i \in N} h^M(x_i)}{|N|}\right) \leq \left(h^M\right)^{-1}\left(q - \frac{1}{M}\right) = \frac{1}{M} < \varepsilon$, so that $\|f(x) - f^q(x)\| < \varepsilon$. ■

Figure 3 depicts the function h^m as described in the proof of Theorem 3.

The quota rule approximation theorem should *not* be interpreted as the statement that quota rules are "approximately" immune to gerrymandering. Rather, the way to read it is that for any quota rule, there exists a rule immune to gerrymandering which pointwise approximates it. The implications of Theorem 3 are that one can design institutions which are not susceptible to gerrymandering, and for which each district "almost" gets its own representative.

Clearly, implementation of such a rule may be difficult. The reason being is that we imagine that all agents care only about the degenerate alternatives; preferences over lotteries can then be extended naturally by stochastic dominance. The reason this is important is that when each district votes, the outcome of a vote is some lottery over the alternatives. The implicit idea is that some representative who votes for this alternative must be sent to some governing body. But if all agents in society prefer one of the alternatives to another, it may be difficult to find an agent who will actually commit to vote for this alternative. One possibility is to send a disinterested individual (an individual who is indifferent between the two alternatives) as a representative. Practically, though, it may be difficult to find a disinterested politician.

4.3 On quasi-proportional representation for non-binary environments

Here, we discuss an environment in which the primitive, underlying set of alternatives is X, which is some arbitrary finite set. Again, we allow agents to submit votes for lotteries over X. Therefore, the domain over which agents vote is now $\Delta(X) \equiv \left\{p \in \mathbb{R}_+^X : \sum_{x \in X} p(x) = 1\right\}$.

All of the axioms previously stated are well-defined in this environment, with the exception of monotonicity, which needs to be reformulated. In the two-alternative case, it was sufficient to say that if every agent voted for a higher value, then in the aggregate, a higher value is selected. In the two-alternative case; however, lotteries are naturally completely ordered. Here, they are not. However; there is still a natural definition of monotonicity that we can discuss.

Thus, let $x \in X$, let $N \in \mathcal{N}$, and let $p, p' \in \Delta(X)$ so that for all $i \in N$, $p_i(x) \geq p'_i(x)$. To extend our idea of monotonicity in the obvious way, a monotonic rule is a rule for which $f(p)(x) \geq f(p')(x)$.

Note that the preceding definition requires that if for all $i \in N$, $p_i(x) = p'_i(x)$, then $f(p)(x) = f(p')(x)$. Hence, the representation of alternative x is a function only of the votes for alternative x; and is independent of the way that voters vote for the other alternatives.

We know from results in the functional equations literature that this type of monotonicity condition is enough to force us to use proportional representation. A nice reference is Ju, Miyagawa, and Sakai (Corollary 10) [10].

Generalized monotonicity For all $N \in \mathcal{N}$ and for all $p, p' \in \Delta(X)^N$ such that for all $x \in X$ and all $i \in N$, $p_i(x) \geq p'_i(x)$, $f(p)(x) \geq f(p')(x)$.

Theorem 4 (Ju, Miyagawa, Sakai): A rule satisfies unanimity, anonymity, and generalized monotonicity if and only if it is the proportional rule.

Note that the preceding theorem does not require representative consistency or continuity. In environments with three or more alternatives, the built-in separability of generalized monotonicity already forces us to use the proportional rule. In two-alternative environments, no such separability is implicit in monotonicity.

The preceding result generalizes a well-known result in the literature on probability aggregation, found in McConway [12]. An immediate consequence is that no result such as the quota rule approximation theorem holds in three-alternative environments. The best we can hope for is to remove generalized monotonicity, but in so doing, we would be forced to admit rules which behave perversely.

4.4 On the spatial model and weakly monotonic voting rules

In models for which there is a natural mixing operation over the set of alternatives, and for which preferences may be single-peaked over those alternatives, there is much more to be said. In such models, the notion of strict monotonicity introduced above is too strong. If we wish agents to always vote for their most preferred alternative, it is clear that strict monotonicity will provide the wrong strategic incentives. To this end, the last question that we address concerns the removal of the strict version of monotonicity discussed in the preceding theorems. The following class of rules is first discussed by Fodor and Marichal [9], although the rules are given a different representation there. However, their work was the first to address the mathematical issue discussed here.

To discuss this question formally, we describe a very general (yet tractable) family of rules. This family is the exhaustive class of rules satisfying several intuitive properties. Recall the use of strict monotonicity in Theorem 2. This section concerns the implications of weakening strict monotonicity.

To begin with, we start with a partition of the unit interval into intervals, say Π. Elements of Π are allowed to be degenerate intervals (singletons). There are several properties that this partition must satisfy. Firstly, as the elements of the partition are themselves intervals, we can order them in the natural way: so that for $\pi, \pi' \in \Pi$, $\pi < \pi'$ if for all $x \in \pi$ and all $x' \in \pi'$, $x < x'$.

We will have occasion to refer to either the left or right endpoint of an interval without specifying whether it is open or closed. The '<' and '>' notation refers to an endpoint of an interval that could be either open or closed.

The partition Π has the feature that for all $\pi \in \Pi$ such that π is right-closed and not a singleton, then for all $\pi' \in \Pi$ such that $\pi' > \pi$, π' is also right-closed. Moreover, for all $\pi \in \Pi$ such that π is left-closed and not a singleton, then for all $\pi' \in \Pi$ such that $\pi' < \pi$, π' is left-closed. Call such a partition an **endpoint connected partition**. The reason for this terminology is simple; a degenerate interval $\{c\}$ can always be viewed the limit of a sequence of half-closed intervals; in this sense, we can interpret $\{c\}$ as a half-closed interval itself, with either a right or left closed endpoint. With this idea, the set of elements of the partition whose right endpoint is closed is an interval according to $<$, and the set of elements of the partition whose left endpoint is closed is an interval according to $<$. In other words, we can classify each singleton as "right-closed" or as "left-closed," so that the set of right-closed and left-closed intervals are connected sets according to $<$.

We give some simple examples to illustrate the concept.

Example 2: Both the trivial partition $\Pi = \{[0,1]\}$ and the finest possible partition $\Pi = \{\{x\} : x \in [0,1]\}$ satisfy the requirements of the definition.

Example 3: Consider the partition $\Pi = \{[1/4, 3/4]\} \cup \{\{x\} : x \notin [1/4, 3/4]\}$. Then this partition is also endpoint connected. This partition is the simple one in which one element is a central interval $[1/4, 3/4]$, and the remaining elements are the singletons which do not lie in $[1/4, 3/4]$.

Example 4: Consider the partition

$$\Pi = \{[0, 1/8), [1/8, 1/4), [1/4, 3/4], (3/4, 7/8], (7/8, 1]\}.$$

This is a 'symmetric' partition, and it can be easily verified that it is endpoint connected.

One feature of endpoint connectedness that is implied by the definition is that there must exist a "central" closed interval, and all other intervals are either half-closed or singletons. Of course, the definition does not imply that there exists a *unique* central closed interval (consider the finest possible partition, as in Example 1), but it does imply that there exists at least one. Singletons can lie either to the right or to the left of the closed interval, but intervals whose right endpoint is closed must be greater than the closed interval, and intervals whose left endpoint is closed must be less than the closed interval. The next example is one in which there is a unique "central" interval, which is the leftmost interval in the partition.

Example 5: Let $\Pi = \{[0, 1/2], (1/2, 1]\}$. In this endpoint connected partition, there are no half-closed intervals whose left endpoint is closed. However, it is clear that this partition satisfies the definition of endpoint connectedness.

The next part of the construction of such a rule is a partial order over the partition. This partial order will have very specific properties. We write $\pi \geq \pi'$ if either $\pi > \pi'$ or $\pi = \pi'$.

Denote by π^* the center closed interval as discussed above. Define the partial order $\succeq$ over Π as follows. For all $\pi, \pi' \geq \pi^*$, if $\pi \leq \pi'$, then $\pi \preceq \pi'$. For all $\pi, \pi' \leq \pi^*$, if $\pi \leq \pi'$, then $\pi \succeq \pi'$. All other pairs remain unordered according to $\succeq$.

The binary relation $\succeq$ should be understood as a "priority" of some intervals over others. It is an example of the notion of "partial priority," which we develop in [3]. The "partial priority" developed on the set of intervals here has the same 'v' shape as those partial priorities which induce target rules, introduced by Thomson [16], and Ching and Thomson [4].

The last part of the construction identifies with each element of the partition a real-valued function. For all $\pi \in \Pi$, define a function $g^{\pi} : \pi \to \mathbb{R}$ which is continuous and strictly increasing. Further, if $\pi = (a, b]$, then $\lim_{x \to a^+} g^{\pi}(x) = -\infty$, and if $\pi = [a, b)$, then $\lim_{x \to a^-} g^{\pi}(a) = \infty$. Of course, if π is a singleton, there is no content to these limiting conditions.

We can now define the rule. The rules we discuss are thus parametrized by three features: an endpoint connected partition of Π, a partial order over this partition, and a collection of functions, one for each element of the partition.

We first discuss intuitively how such a rule works. Informally, the way the rule aggregates votes is as follows. For every possible vote profile, we can find a unique "highest priority" interval containing a vote in the following way. Each interval which contains a vote is put into a set. We can then take the meet of these intervals according to the partial order (this meet is well-defined by construction). This meet is the lowest priority interval having a priority weakly higher than all intervals receiving votes.

By construction, if this interval lies weakly to the right of all intervals receiving votes, then it has a closed-left endpoint. We then replace all votes to the left of this endpoint with the endpoint. Analogously, if the interval lies weakly to the left of all intervals receiving votes, then it has a closed right endpoint. We then replace all votes to the right of this endpoint with the endpoint. Lastly, if it lies somewhere in the center of all intervals receiving votes, it must itself be the closed center interval. In this case, any votes to the right of the interval are mapped to the right endpoint and any votes to the left are mapped to the left endpoint.

We then obtain a revised profile of votes which lie completely in one of the intervals. Associated with the interval is a function. We use this function to take the quasi-arithmetic mean of the revised votes. The value of this mean is then the outcome of the vote.

Formally, let $N \in \mathcal{N}$ and $x \in [0,1]^N$. For all $x_i \in [0,1]$, define $\pi(x_i) \in \Pi$ to be the element of the partition in which x_i lies; so that $\pi(x_i) \equiv \{\pi \in \Pi : x_i \in \pi\}$. For any element $\pi \in \Pi$, π may be written $\pi = \langle a^{\pi}, b^{\pi} \rangle$, for some endpoints a^{π} and b^{π} (if π is degenerate, then $a^{\pi} = b^{\pi}$). For all $y \in [0,1]$, define $y^{\pi} \in [0,1] = med\{a^{\pi}, b^{\pi}, y\}$. Thus, y^{π} is the 'closest' point in the closure of the interval to y. Therefore, $\bigwedge\limits_{i \in N} \pi(x_i)$ is the meet of the intervals receiving votes according to the partial order.

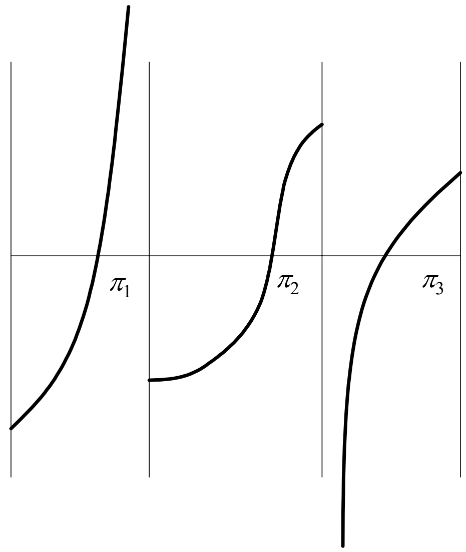

Figure 4: A generalized target rule

We define $f(x) = \left(g^{\bigwedge_{i\in N} \pi(x_i)} \right)^{-1} \left(\frac{\sum_{i\in N} g^{\bigwedge_{i\in N} \pi(x_i)} \left(x_i^{\bigwedge_{i\in N} \pi(x_i)} \right)}{|N|} \right)$.

The preceding expression generates what we call the generalized target rule with respect to $\left(\Pi, \left(\preceq, \{g^\pi\}_{\pi\in\Pi}\right)\right)$. The expression is quite involved, and needs to be explained. We will do this with several simple examples. But first, we present a figure.

In the diagram, the endpoint connected partition consists of $\Pi = \{\pi_1, \pi_2, \pi_3\}$. The center, closed, interval is π_2. The function g^{π_2} associated with π_2 is thus continuous and strictly increasing. The interval π_1 is open to the right; indeed, g^{π_1} approaches infinity as x tends to the right endpoint of π_1. Likewise, the interval π_3 is open to the left, so that g^{π_3} approaches infinity as x tends to this left endpoint of π_3.

Example 6: A standard target rule obtains when Π is the finest possible partition; i.e. $\Pi = \{\{x\} : x \in [0,1]\}$. The functions g^π are allowed to be anything. The "target" is defined simply by setting a to be the "center" interval described above (in other words, the alternative a such that for all $x \in [0,1]$, $a \preceq x$). This rule is the rule which, for any vote profile x, selects that alternative in the interval spanned by the agents' votes which lies closest to the "target."

Example 7: A rule which is a quasi-arithmetic mean results when $\Pi = \{[0,1]\}$ and

$g^{[0,1]}$ is some strictly increasing, continuous function. This rule always selects $f(x) = \left(g^{[0,1]}\right)^{-1}\left(\frac{\sum_{i\in N} g^{[0,1]}(x_i)}{|N|}\right)$. Such rules were axiomatized by Kolmogorov, Nagumo, and Aczél [1]. These are discussed in Theorem 2 above.

Example 8: Another example is that of the geometric mean: This is the rule which is defined by $f(x) = \sqrt[|N|]{\prod_{i\in N} x_i}$. This rule results when the partition is given by $\Pi = \{\{0\}, (0,1]\}$. Of course, $g^{\{0\}}$ can be arbitrary, but here, $g^{(0,1]}(x) = \log(x)$. The element $\{0\}$ is the "center" interval referred to above. Note that, as required, $\lim_{x\to 0^+} \log(x) = -\infty$. Moreover, $\{0\} \preceq (0,1]$. Thus, any time that $\{0\}$ receives any votes, it is the selected alternative. In a sense, $\{0\}$ has a "priority" over all other alternatives.

Example 9: A "centralizing rule." Let $\Pi = \{[1/4,3/4]\} \cup \{\{x\} : x \notin [1/4,3/4]\}$. Thus, Π is a partition into one non-degenerate center interval $[1/4,3/4]$, where the remaining elements of the partition are degenerate singletons. Define $g^{[1/4,3/4]}(x) = x$, and for all remaining elements π, g^{π} is arbitrary. The partial order ranks $[1/4,3/4] \preceq \pi$ for all intervals, and for degenerate intervals to the right of $[1/4,3/4]$, say $\{x\}$ and $\{y\}$, $\{x\} \preceq \{y\}$ if and only if $x \leq y$. For degenerate intervals to the left of $[1/4,3/4]$, $\{x\} \preceq \{y\}$ if and only if $x \geq y$. This rule works as follows. For all $N \in \mathcal{N}$, for all $x \in [0,1]^N$, if for all $i \in N$, $x_i \leq 1/4$, then $f(x) = \max_{i\in N} x_i$, and if for all $i \in N$, $x_i \geq 3/4$, then $f(x) = \min_{i\in N} x_i$. Otherwise, $f(x) = \frac{\sum_{i\in N} med\{1/4,3/4,x_i\}}{|N|}$. In other words, if every voter votes for an extreme right alternative (so that $x_i \geq 3/4$), then the rule selects the least right of the alternatives. Likewise, if every agent votes for an extreme left alternative, then the rule selects the least left of the alternatives. In every other case, those voting for extreme right alternatives are treated as if they had voted for $3/4$, and those voting for extreme left alternatives are treated as if they had voted for $1/4$. Proportional representation is then applied. Thus, extreme votes are moderated before being aggregated by proportional representation.

A characterization of this class of rules is provided in the mathematics literature, by Fodor and Marichal [9]. Thus, they provide the first full description of this family. Their representation is of a different type, which is more difficult to interpret in terms of political environments. Specifically, their representation does not discuss the idea of an "endpoint-connected partition," or of a priority ordering over different segments of the unit interval. In contrast, the representation we discuss here emphasizes the priority-like nature of the set of intervals in the partition, and connects the class in an intuitive way with the class characterized for finite alternative environments, discussed in [3]. However, it should be clear that the class previously characterized by Fodor and Marichal is the same as the class we characterize here (as it must be). Indeed, their proof shares many characteristics with the one provided here. A proof is given for the sake of completeness, and some parts follow Fodor and Marichal.

Theorem 5: A rule satisfies unanimity, anonymity, representative consistency, monotonicity, and continuity if and only if it is a generalized target rule.

To interpret these rules, we suggest the idea of a hierarchy of intervals. The central, closed interval is the most "moderate" of all, and the intervals are ranked according to how moderate they are. Thus, the closer an interval is to the center, the more moderate it is. Such a rule will not let the outcome of a vote be "too extreme," unless everyone in society votes for something extreme.

The rule works in different ways, depending on the position of the votes. But suppose, for example, that there are some individuals who vote for alternatives to the right of the moderate interval, and some who vote for alternatives to the left of the moderate interval. Those votes for extreme alternatives (alternatives outside of the center interval) are then mapped to their more moderate counterparts (the closest points in the moderate interval to these votes). The outcome of the vote is then within a moderate region. Such a system could theoretically be used to prevent the outcome of a vote from being too "extreme," where extreme is defined in some absolute sense. It allows all voters some degree of veto power over extreme alternatives. Note that, in a sense, the wider are the intervals, the greater the opportunity for strategic manipulation on the part of voters.

Of course, the notion of which intervals are ranked and which interval is the center interval is completely undetermined by our axioms. Hence; notions of moderation and of extremism must be subjective. This flexibility is an obvious benefit in an abstract model, for which the interpretation of elements of $[0,1]$ is left completely arbitrary. In more concrete applications, the structure and interpretation of elements of $[0,1]$ should be used in guiding a decision as to what the hierarchy of intervals should be.

5 Conclusion

In this work, we show how to construct consistent systems of representation by allowing voters to vote for the composition of a governing body. Such constructs allow us to bypass the impossibility result of [3]. In the case of binary social decisions, one can construct rules which approximate any quota rule (to an arbitrarily high degree) and are immune to the phenomenon of gerrymandering.

We also study this issue in non-binary environments. We show that in such environments, the proportional rule is essentially the only natural (*i.e.* monotonic) rule available. Thus, the possibilities in this environment are much more restricted.

We discuss the distinction between monotonicity and strict monotonicity. Dropping the requirement that rules be strictly monotonic allows us to introduce many more rules. These rules are compatible with the idea of an exogenous, moderate set of alternatives. Any voter in society can veto alternatives which are too extreme (by forcing the outcome of a vote to lie in the moderate set of alternatives).

6 Appendix

Here, we offer a proof of Theorem 5. The proof relies on classical results from the theory of functional equations. For two alternatives x, y, the notation $x \preceq y$ means that $f(x, y) = x$.

Lemma 1: Let f satisfy our main axioms. Suppose that $x < y$, and that $f(x, y) = x$. Then for all $z > x$, $f(x, z) = x$.

Proof. First, suppose that $z \in (x, y)$. By monotonicity, $f(x, x) \leq f(x, z) \leq f(x, y)$. Therefore, $f(x, z) = x$.

Next, let $z^* \equiv \sup\{z : f(x, z) = x\}$. By continuity, $f(x, z^*) = x$. Moreover, by the preceding step, for all $z \in (x, z^*)$, $f(x, z) = x$. We now show that $z^* = 1$. Suppose, by means of contradiction, that $z^* < 1$. Then as $x < y$, by continuity, there exists $\varepsilon > 0$ so that $z^* + \varepsilon < 1$ and $f(x, z^* + \varepsilon) < y$. By monotonicity, $x = f(x, x) \leq f(x, z^* + \varepsilon)$. Label $w \equiv f(x, z^* + \varepsilon)$.

By replication invariance, $f(x, z^* + \varepsilon) = f(x, x, z^* + \varepsilon, z^* + \varepsilon)$. By representative consistency, $f(x, x, z^* + \varepsilon, z^* + \varepsilon) = f(x, w, w, z^* + \varepsilon)$. By representative consistency, $f(x, w, w, z^* + \varepsilon) = f(f(x, w), f(x, w), w, z^* + \varepsilon)$. As $x \leq w < y$, $f(x, w) = x$, so that $f(f(x, w), f(x, w), w, z^* + \varepsilon) = f(x, x, w, z^* + \varepsilon)$. By representative consistency, $f(x, x, w, z^* + \varepsilon) = f(x, f(x, w), f(x, w), z^* + \varepsilon) = f(x, x, x, z^* + \varepsilon)$. By representative consistency, $f(x, x, x, z^* + \varepsilon) = f(x, x, w, w)$. By replication invariance, $f(x, x, w, w) = f(x, w)$. But $f(x, w) = x$. Hence $f(x, z^* + \varepsilon) = x$, contradicting the definition of z^*. Hence for all $z > y$, $f(x, y) = x$. ■

Lemma 1 is symmetric (by reversing the inequalities, we maintain the statements of the lemmas). Having established the preceding, we define the set of right endpoints, $\mathcal{R} \equiv \{x : f(x, 1) = x\}$ and left endpoints, $\mathcal{L} \equiv \{x : f(0, x) = x\}$. Each of these sets is clearly closed (by continuity). We also claim that for all $x \in \mathcal{R}$ and all $y \in \mathcal{L}$, $y \leq x$. To see this, suppose $x \in \mathcal{R}$ and $y \in \mathcal{L}$. Suppose that $y > x$. As $x \in \mathcal{R}$, $f(x, 1) = x$. As $y > x$, by Lemma 1, $f(x, y) = x$. Moreover, as $y \in \mathcal{L}$, $f(0, y) = y$, and as $x < y$, by Lemma 1, $f(x, y) = y$. Thus, $y = f(x, y) = x$. This contradicts $y > x$.

Let Π be the partition of $[0, 1]$ constructed accordingly, so that $\mathcal{R}$ is the set of right endpoints and $\mathcal{L}$ is the set of left endpoints. $\mathcal{R}$ and $\mathcal{L}$ can have at most one point in common (by the preceding), so that we get an endpoint-connected partition (as described in the main theorem).

The next lemma establishes a simple fact that will be useful in the remainder of the proof.

Lemma 2: Let $N \in \mathcal{N}$ and $x \in X^N$. Let $y \in X$. Suppose that $f(x, y) = f(x)$. Then $f(x) \preceq y$.

Proof. Suppose the hypotheses of the lemma are satisfied, and label $|N| = n$. By representative consistency, $f(x,y) = f(f(x)^n, y)$. Therefore, $f(f(x)^n, y) = f(x)$. Now, consider $f\left(f(x)^{n+1}, y\right)$. By anonymity, we may rewrite $f\left(f(x)^{n+1}, y\right) = f(f(x)^n, y, f(x))$. By representative consistency, $f(f(x)^n, y, f(x)) = f\left(f(f(x)^n, y)^{n+1}, f(x)\right)$. But as $f(f(x)^n, y) = f(x)$, the preceding is equal to $f\left(f(x)^{n+1}, f(x)\right)$. By unanimity, conclude $f\left(f(x)^{n+1}, f(x)\right) = f(x)$. Therefore, $f\left(f(x)^{n+1}, y\right) = f(x)$ and $f(f(x)^n, y) = f(x)$. We may thus conclude from Theorem 1 that $f(f(x), y) = f(x)$, so that $f(x) \preceq y$. ■

Lemma 3: For all $\pi \in \Pi$, f restricted to π is strictly monotonic.

Proof. To show this, it is enough to let $N \in \mathcal{N}$, $x \in \pi^N$ and $y, z \in \pi$ such that $y < z$. Suppose $|N| = n$. We will show that $f(x,y) < f(x,z)$. Suppose, by means of contradiction, that $f(x,y) = f(x,z)$. We will derive a contradiction in several cases.

Suppose first that $y \leq f(x) \leq z$. Then by monotonicity, $y \leq f(f(x)^n, y) \leq f(x) \leq f(f(x)^n, z) \leq z$. But by representative consistency, $f(f(x)^n, y) = f(x,y)$ and $f(f(x)^n, z) = f(x,z)$. Thus, as $f(x,y) = f(x,z)$, $f(x,y) = f(x) = f(x,z)$. If $y = f(x)$, then $f(x) < z$ and $f(x,z) = f(x)$. By Lemma 2, $f(f(x), z) = f(x)$. By Lemma 1, $f(f(x), z) = f(x)$ and $f(x) < z$ implies that $f(x) \in \mathcal{R}$, in which case there exists $i \in N$ such that $x_i \leq f(x) < z$, contradicting the fact that x_i and z are in the same interval. In the case that $y < f(x)$, we may similarly conclude $f(x) \in \mathcal{L}$. But then there exists $x_i \geq f(x) > y$, contradicting the fact that x_i and y are in the same interval.

Next, suppose that $f(x) \leq y < z$. If $y = f(x)$, then $f(x,y) = f(f(x)^n, y) = f(x)$, where the last equality follows from unanimity. Conclude that $f(x,z) = f(x)$. By Lemma 2, $f(f(x), z) = f(x) < z$, so that $f(x) \in \mathcal{R}$, establishing a contradiction. Thus, we may assume without loss of generality that $f(x) < y < z$.

Define $w \equiv \inf\{c : f(x,c) = f(x,z)\}$. Then by continuity, $f(x,w) = f(x,z)$. In particular, $w \leq y < z$. We claim that either $w \in \mathcal{R}$ or $f(x,w) \in \mathcal{L}$. By the argument in the preceding paragraph, we may assume without loss of generality that $w > f(x)$.

Suppose that $w \notin \mathcal{R}$. Then by definition of $\mathcal{R}$, $w < f(w,z) \leq z$. As f is continuous, there exists $\varepsilon > 0$ small so that $w < f(w-\varepsilon, z) \leq z$. Thus, by monotonicity, $f(x, f(w-\varepsilon, z)) = f(x,w)$. By replication invariance, $f(x, f(w-\varepsilon, z)) = f(x, x, f(w-\varepsilon, z), f(w-\varepsilon, z))$. By representative consistency, $f(x, x, f(w-\varepsilon, z), f(w-\varepsilon, z)) = f(x, x, w-\varepsilon, z)$. By anonymity, $f(x, x, w-\varepsilon, z) = f(x, z, x, w-\varepsilon)$. By representative consistency, $f(x, z, x, w-\varepsilon) = f\left(f(x,z)^{n+1}, f(x, w-\varepsilon)^{n+1}\right)$. By replication invariance, $f\left(f(x,z)^{n+1}, f(x, w-\varepsilon)^{n+1}\right) = f(f(x,z), f(x, w-\varepsilon))$. As $f(x,z) = f(x,w)$, $f(f(x,z), f(x,w-\varepsilon)) = f(f(x,w), f(x,w-\varepsilon))$. Lastly, by definition of w, $f(x, w-\varepsilon) < f(x,w)$. Therefore, stringing together the equalities,

$f(f(x,w), f(x,w-\varepsilon)) = f(x,w)$, so that $f(x,w) \in \mathcal{L}$. Moreover, it is clear that $f(x) < f(x,w)$. This follows because $f(x) \leq f(x,w-\varepsilon)$ (for ε small enough), so that $f(x) \leq f(x,w-\varepsilon) < f(x,w)$. In this case, conclude that $f(x)$ does not lie in the same interval as w and z (and hence y and z), so that there exists some x_i not lying in the same interval as y and z.

If, in fact $w \in \mathcal{R}$, then z does not lie in the same interval as $f(x)$, and hence some x_i.

The remaining cases are symmetric to the ones presented above. Therefore the conclusion of the lemma holds. ■

Lemma 3 establishes that restricted to any $\pi \in \Pi$, f satisfies the axioms necessary for representation as a quasi-arithmetic mean. Thus, with each element $\pi \in \Pi$, there exists a function (continuous and strictly increasing) $g^\pi : \pi \to \mathbb{R}$ such that for all $N \in \mathcal{N}$ and all $x \in \pi^N$, $f(x) = (g^\pi)^{-1}\left(\frac{\sum_N g^\pi(x_i)}{|N|}\right)$.

Suppose that the interval π is open on the left, so that $\pi = (a^\pi, b^\pi]$. Then we claim that $\lim_{x \to a^\pi +} g^\pi(x) = -\infty$. Suppose, by means of contradiction, that the statement is false. The right-hand limit exists as g^π is continuous and strictly increasing; let us thus extend g^π to $[a^\pi, b^\pi]$ so that $g^\pi(a^\pi) = \lim_{x \to a^\pi +} g^\pi(x)$. The continuity of g^π is therefore preserved.

Now, as $a^\pi \in \mathcal{R}$, it follows from Lemma 1 that for all $y \in \pi$, $f(a^\pi, y) = a^\pi$. Let $\{x_n\}_{n=1}^{\infty} \subset \pi$ such that $\lim x_n = a^\pi$. By continuity, $f(x_n, y) \to f(a^\pi, y) = a^\pi$. But $f(x_n, y) = (g^\pi)^{-1}\left(\frac{g(x_n)+g(y)}{2}\right)$. Moreover, as $f(x_n, y) \to a^\pi$, by monotonicity, we conclude that $f(x_n, y) = a^\pi + \varepsilon_n$, for some $\varepsilon_n \to 0$, $\varepsilon_n \geq 0$. Therefore, $\frac{g^\pi(x_n)+g^\pi(y)}{2} = g^\pi(a^\pi + \varepsilon_n)$. By letting $n \to \infty$, we obtain $\frac{g^\pi(a^\pi)+g^\pi(y)}{2} = g^\pi(a^\pi)$. Solving obtains $g^\pi(a^\pi) = g^\pi(y)$. But this contradicts the fact that g^π is strictly increasing. Therefore, $\lim_{x \to a^\pi +} g^\pi(x) = -\infty$. A similar statement shows that if π is open to the right, then $\lim_{x \to b^\pi -} g^\pi(x) = +\infty$.

The last step of the proof is to establish that f is a generalized target rule with partition Π and functions $\{g^\pi\}_{\pi \in \Pi}$. In order to establish this, we need a few more results. This part of the proof follows the work of Fodor and Marichal [9].

Thus, let $\pi = (a^\pi, b^\pi]$ be any interval with a right-closed, left-open endpoint. Let $x \in \pi$, and suppose that $y \geq b^\pi$. Then we claim that $f(x,y) = f(x, b^\pi)$. To establish this, $f(x, b^\pi) = f(x, x, b^\pi, b^\pi)$ by replication invariance. Now, from the representation of f on π as given by g^π, we conclude that there exists some $z \in (a^\pi, b^\pi]$ such that $f(z, b^\pi) = x$. Therefore, $f(x, x, b^\pi, b^\pi) = f(f(z, b^\pi), f(z, b^\pi), b^\pi, b^\pi)$. By representative consistency, $f(f(z, b^\pi), f(z, b^\pi), b^\pi, b^\pi) = f(z, b^\pi, b^\pi, b^\pi)$. But, as $b^\pi \in \mathcal{R}$, $f(b^\pi, y) = b^\pi$, so that $f(z, b^\pi, b^\pi, b^\pi) = f(z, b^\pi, f(b^\pi, y), f(b^\pi, y))$. By representative consistency, $f(z, b^\pi, f(b^\pi, y), f(b^\pi, y)) = f(z, b^\pi, b^\pi, y)$. Again, $f(z, b^\pi, b^\pi, y) = f(z, f(b^\pi, y), f(b^\pi, y), y)$. By representative consistency,

$f(z, f(b^\pi, y), f(b^\pi, y), y) = f(z, b^\pi, y, y)$. Again by representative consistency, $f(z, b^\pi, y, y) = f(f(z, b^\pi), f(z, b^\pi), y, y)$. But $f(z, b^\pi) = x$, so that $f(f(z, b^\pi), f(z, b^\pi), y, y) = f(x, x, y, y)$. By replication invariance, $f(x, x, y, y) = f(x, y)$. Therefore, $f(x, y) = f(x, b^\pi)$. A similar argument verifies that if $\pi = [a^\pi, b^\pi)$ and $x \in \pi$ and $z \leq a^\pi$, then $f(x, y) = f(x, a^\pi)$.

Next, consider the interval $\pi = [a^\pi, b^\pi]$. This is the center closed interval as described in the definition of generalized target rule. The point a^π is the right-most point in $\mathcal{L}$ and the point b^π is the left-most point in $\mathcal{R}$. (These right-most and left-most points clearly exist, as $\mathcal{L}$ and $\mathcal{R}$ are closed sets, by continuity of f).

We first establish that for all $x \leq a^\pi$ and for all $y \in [a^\pi, f(a^\pi, 1)]$, $f(x, y) = f(a^\pi, y)$. Thus, as $y \in [a^\pi, f(a^\pi, 1)]$, and as $f(a^\pi, a^\pi) = a^\pi$, there exists by continuity some $z \in [a^\pi, 1]$ such that $y = f(a^\pi, z)$. Then $f(x, y) = f(x, f(a^\pi, z))$. By replication invariance, $f(x, f(a^\pi, z)) = f(x, x, f(a^\pi, z), f(a^\pi, z))$. By representative consistency, $f(x, x, f(a^\pi, z), f(a^\pi, z)) = f(x, x, a^\pi, z)$. By representative consistency, $f(x, x, a^\pi, z) = f(f(x, a^\pi), f(x, a^\pi), f(x, z), f(x, z))$. As $x \leq a^\pi$ and as $a^\pi \in \mathcal{L}$, $f(x, a^\pi) = a^\pi$. Hence $f(f(x, a^\pi), f(x, a^\pi), f(x, z), f(x, z)) = f(a^\pi, a^\pi, f(x, z), f(x, z))$. By representative consistency, $f(a^\pi, a^\pi, f(x, z), f(x, z)) = f(a^\pi, a^\pi, x, z)$. By representative consistency, $f(a^\pi, a^\pi, x, z) = f(f(a^\pi, x), f(a^\pi, x), f(a^\pi, z), f(a^\pi, z))$. But the preceding is equal to $f(a^\pi, a^\pi, y, y)$. By replication invariance, $f(a^\pi, a^\pi, y, y) = f(a^\pi, y)$. Therefore, $f(x, y) = f(a^\pi, y)$.

A symmetric argument establishes that for all $y \geq b^\pi$ and for all $x \in [f(0, b^\pi), b^\pi]$, $f(x, y) = f(x, b^\pi)$.

Next, we claim that if $x \leq a^\pi$ and $y \geq b^\pi$, then $f(x, y) = f(a^\pi, b^\pi)$. To see why, by representative consistency, $f(x, y) = f(f(x, y), f(x, y))$. By replication invariance, $f(f(x, y), f(x, y)) = f(f(x, y), f(x, y), f(x, y), f(x, y))$. By representataive consistency, $f(f(x, y), f(x, y), f(x, y), f(x, y)) = f(x, y, f(x, y), f(x, y))$. By representative consistency, $f(x, y, f(x, y), f(x, y)) = f(f(x, f(x, y)), f(x, f(x, y)), f(y, f(x, y)), f(y, f(x, y)))$. Now, $f(x, y) \in [a^\pi, f(a^\pi, 1)]$. To see this, note that $f(x, y) \geq f(x, a^\pi)$ by monotonicity, and since $f(x, a^\pi) = a^\pi$, we obtain $f(x, y) \geq a^\pi$. Moreover, as $x \leq a^\pi$ and $y \leq 1$, $f(x, y) \leq f(a^\pi, 1)$. Conclude by the previous paragraph that $f(x, f(x, y)) = f(a^\pi, f(x, y))$. Similarly, we may conclude that $f(y, f(x, y)) = f(b^\pi, f(x, y))$. Therefore, $f(f(x, f(x, y)), f(x, f(x, y)), f(y, f(x, y)), f(y, f(x, y))) = f(f(a^\pi, f(x, y)), f(a^\pi, f(x, y)), f(b^\pi, f(x, y)), f(b^\pi, f(x, y)))$. By representative consistency, $f(f(a^\pi, f(x, y)), f(a^\pi, f(x, y)), f(b^\pi, f(x, y)), f(b^\pi, f(x, y))) = f(a^\pi, f(x, y), b^\pi, f(x, y))$. By representative consistency, $f(a^\pi, f(x, y), b^\pi, f(x, y)) = f(x, y, a^\pi, b^\pi)$. By representative consistency, $f(x, y, a^\pi, b^\pi) = f(f(x, a^\pi), f(x, a^\pi), f(y, b^\pi), f(y, b^\pi))$. But as $f(x, a^\pi) = a^\pi$ and as $f(y, b^\pi) = b^\pi$, we conclude that $f(f(x, a^\pi), f(x, a^\pi), f(y, b^\pi), f(y, b^\pi)) = f(a^\pi, a^\pi, b^\pi, b^\pi)$. Finally, by replication invariance, $f(a^\pi, a^\pi, b^\pi, b^\pi) = f(a^\pi, b^\pi)$. Hence, $f(x, y) = f(a^\pi, b^\pi)$ in this case.

Lastly, we show that if $x \leq a^{\pi}$ and $y \in [a^{\pi}, b^{\pi}]$, then $f(x,y) = f(a^{\pi}, y)$. A symmetric argument establishes that if $y \geq b^{\pi}$ and $x \in [a^{\pi}, b^{\pi}]$, then $f(x,y) = f(x, b^{\pi})$. We have already established that the statement holds for all $y \leq f(a^{\pi}, 1)$. Thus, suppose that $y \geq f(a^{\pi}, 1)$. By the previous argument, this is equivalent to the statement that $y \geq f(a^{\pi}, b^{\pi})$. Thus, as $f(b^{\pi}, b^{\pi}) = b^{\pi}$, by continuity, there exists some $z \in [a^{\pi}, b^{\pi}]$ such that $y = f(b^{\pi}, z)$. Now, $f(x,y) = f(x, f(b^{\pi}, z))$. By replication invariance, $f(x, f(b^{\pi}, z)) = f(x, x, f(b^{\pi}, z), f(b^{\pi}, z))$. By representative consistency, $f(x, x, f(b^{\pi}, z), f(b^{\pi}, z)) = f(x, x, b^{\pi}, z)$. By representative consistency, $f(x, x, b^{\pi}, z) = f(f(x, b^{\pi}), f(x, b^{\pi}), f(x,z), f(x,z))$. By the preceding step, $f(x, b^{\pi}) = f(a^{\pi}, b^{\pi})$. Therefore, $f(f(x, b^{\pi}), f(x, b^{\pi}), f(x,z), f(x,z)) = f(f(a^{\pi}, b^{\pi}), f(a^{\pi}, b^{\pi}), f(x,z), f(x,z))$. By representative consistency, $f(f(a^{\pi}, b^{\pi}), f(a^{\pi}, b^{\pi}), f(x,z), f(x,z)) = f(a^{\pi}, b^{\pi}, x, z)$. By representative consistency, $f(a^{\pi}, b^{\pi}, x, z) = f(f(x, a^{\pi}), f(x, a^{\pi}), f(b^{\pi}, z), f(b^{\pi}, z))$. But as $f(x, a^{\pi}) = a^{\pi}$ and $f(b^{\pi}, z) = y$, we conclude $f(f(x, a^{\pi}), f(x, a^{\pi}), f(b^{\pi}, z), f(b^{\pi}, z)) = f(a^{\pi}, a^{\pi}, y, y)$. By replication invariance, $f(a^{\pi}, a^{\pi}, y, y) = f(a^{\pi}, y)$, so that $f(x,y) = f(a^{\pi}, y)$.

To conclude the proof, we verify that f is a generalized target rule with partition Π and functions $\{g^{\pi}\}_{\pi \in \Pi}$. Thus, let $N \in \mathcal{N}$ and $x \in X^N$. Fix $\bigwedge_{i \in N} \pi(x_i)$, where the meet is defined with respect to the partial order discussed in the definition of generalized target rule. Suppose that $x_i > b^{\bigwedge_{i \in N} \pi(x_i)}$ (so that it lies to the right of the meet of the elements). Then we claim that $f\left(x_{N \setminus \{i\}}, x_i\right) = f\left(x_{N \setminus \{i\}}, b^{\bigwedge_{i \in N} \pi(x_i)}\right)$. But this is trivial; by definition of $\bigwedge_{i \in N} \pi(x_i)$, there exists some $j \in N$ such that $x_j \leq b^{\bigwedge_{i \in N} \pi(x_i)}$; moreover, $b^{\bigwedge_{i \in N} \pi(x_i)} \in \mathcal{R}$. Therefore, $f\left(x_{N \setminus \{i.j\}}, x_i, x_j\right) = f\left(x_{N \setminus \{i,j\}}, f(x_i, x_j), f(x_i, x_j)\right)$ by representative consistency. By Lemma 2 and monotonicity, $f(x_i, x_j) = f\left(b^{\bigwedge_{i \in N} \pi(x_i)}, x_j\right)$. Therefore,

$$f\left(x_{N \setminus \{i,j\}}, f(x_i, x_j), f(x_i, x_j)\right) = f\left(x_{N \setminus \{i,j\}}, f\left(b^{\bigwedge_{i \in N} \pi(x_i)}, x_j\right), f\left(b^{\bigwedge_{i \in N} \pi(x_i)}, x_j\right)\right).$$

But the preceding is equal to $f\left(x_{N \setminus \{i\}}, b^{\bigwedge_{i \in N} \pi(x_i)}\right)$ by representative consistency. Proceeding in this fashion, we establish that $f(x) = f\left(\left(x_i^{\bigwedge_{i \in N} \pi(x_i)}\right)_{i \in N}\right)$, as given in the definition of a generalized target rule. But clearly, for all $i \in N$, $x_i^{\bigwedge_{i\in N}\pi(x_i)} \in \bigwedge_{i\in N} \pi(x_i)$. We may therefore apply the quasi-arithmetic mean associated with interval π to obtain that $f(x) = \left(g^{\bigwedge_{i\in N}\pi(x_i)}\right)^{-1}\left(\frac{\sum_{i\in N} g^{\bigwedge_{i\in N}\pi(x_i)}\left(x_i^{\bigwedge_{i\in N}\pi(x_i)}\right)}{|N|}\right)$, concluding the proof. ■

References

[1] J. Aczél, "Lectures on Functional Equations and Their Applications," Academic Press Inc. New York, NY, (1966).

[2] M. Balinski and H.P. Young, "Fair Representation: Meeting the Ideal of One Man, One Vote," 2nd. edition, Brookings Institution Press, Washington D.C., (2001).

[3] C.P. Chambers, "Consistent Representative Democracy," manuscript, (2003).

[4] S. Ching and W. Thomson, "Population-monotonic solutions in public good economies with single-peaked preferences," mimeo, University of Rochester, (1993).

[5] K. Fine, Some necessary and sufficient conditions for representative democracy on two alternatives, *Econometrica* **40** (1972), 1083-1090.

[6] B. De Finetti, Sul concetto di media, *Giornale dell' Instituto Italiano degli Attuari* **2** (1931), 369-396.

[7] P.C. Fishburn, The theory of representative majority decision, *Econometrica* **39** (1971), 273-284.

[8] P.C. Fishburn, "The Theory of Social Choice," Princeton University Press, Princeton, NJ, (1973).

[9] J. Fodor and J.-L. Marichal, On nonstrict means, *Aequationes Mathematicae* **54** (1997), 308-327.

[10] B.-G. Ju, E. Miyagawa, and T. Sakai, "Non-manipulable division rules in claims problems and generalizations," manuscript, (2003).

[11] A. Kolmogorov, (1930), Sur la notion de la moyenne, *Rendiconti Accademia dei Lincei (6)* **12** (1930), 388-391.

[12] K.J. McConway, Marginalization and linear opinion pools, *Journal of the American Statistical Association* **76** (1981), 410-414.

[13] Y. Murakami, Formal structure of majority decision, *Econometrica* **34** (1966), 709-718.

[14] Y. Murakami, "Logic and Social Choice," Dover Publications, New York, NY, (1968).

[15] M. Nagumo, Uber eine Klasse der Mittelwerte, *Japan Journal of Mathematics* **7** (1930), 71-79.

[16] W. Thomson, The replacement principle in public good economies with single-peaked preferences, *Economics Letters* **42** (1993), 31-36.