Choice-based Measures of Conflict in Preferences

Katherine A. Baldiga and Jerry R. Green
Harvard University

March 2, 2010

Abstract

We propose a family of measures of difference between ordinal preference relations. The difference between two preferences is the probability that they would disagree about the optimal choice from a random available set. It is in this sense that these measures are choice-based. Measures differ according to the distribution of the random available sets. We use these measures to propose new social choice rules that achieve maximal expected assent among the members of the population. We also propose two further applications of these measures. The first is to welfare measurement when choice is irrational. The second is to the measurement of polarization in a population.

1 Introduction

We define a family of metrics on the space of ordinal preference relations and trace the implications of these metrics for social choice theory and other problems. Our metrics are choice-based in that they depend on the distribution of the actual feasible set – a member of the collection of all subsets of alternatives. The main idea is that social choice, as a function of the preferences of a group of individuals, should depend on the relative frequency of the decision problems that these individuals might face. Doubt about the actual feasible set is one justification for Arrow's [3] requirement that social choice be derivable from an ordering. Through that requirement Arrow was able to insure that the choice would display collective rationality whatever the feasible set turned out to be. Thus, in our choice-based methodology, the distribution on the feasible sets determines the metric on preferences, and from this metric and a knowledge of the population of preferences we proceed to various economic applications.

Our metrics are simple to explain. Given a distribution over the feasible sets the distance between any two preferences is the probability that these preference would disagree about the optimal decision. Two people will agree on the optimal choice if the differences in their preferences involve only unavailable alternatives or alternatives that they both rank lower than those alternatives that they do agree on. Differences between the ordinal preferences with respect to highly-ranked alternatives will produce more conflict than differences between lower-ranked alternatives because there will be more feasible sets at which such high-ranked differences will matter.

In applying these ideas to social choice theory we evaluate a potential social ordering by computing the probability that a randomly-selected member of the population agrees with the choice that it would induce. The social ordering we propose is the one that maximizes expected agreement, or assent, to the social choice. We show how the assent-maximizing ordering depends upon the statistical assumptions that are made about the distribution of feasible alternatives.

We relate the assent-maximizing social orderings to the one discussed by Kemeny [13], Kemeny and Snell [14] and Young [19], as axiomatized by Young and Levenglick [20]. Their solution, which we call the Kemeny Rule, is the assent maximizing rule when the feasible set is sure to be a pair of alternatives and all pairs are equally likely. The Kemeny Rule is one of our social welfare functions, and as is well-known, the associate metric is Kendall's τ. However whenever the distribution over feasible sets allows for sets larger than pairs, our metric will differ from Kendall's τ and the social welfare function will differ from Kemeny's Rule.

Kemeny, Kemeny-Snell and Young-Levenglick were motivated by Condorcet's wish to "break" cycles in pairs in the least intrusive way possible, reversing pairwise choices that are supported by weak majorities. In this paper, we construct an ordering that aims to respect choices from all subsets, to the maximum extent possible. This ordering reverses majority choices on some pairs in order to better match the the population's choices on subsets of larger cardinalities. We are motivated by the decision-theoretic idea that the feasible set could well be a large subset of the set of all alternatives, and thus we adopt the objective of staying as close as possible to the individuals' wishes on these sets. Pairs are treated in the same way as feasible sets of any other cardinality.

In addition to social choice theory, we give two further applications of this family of metrics. The first is to the problem of multiple-self explanations of irrational individual choice functions.[1] Here the question is how to make a good selection from a set of alternative multi-self explanations of an irrational choice function.

The second application is to measurement of polarization in a society. Esteban and Ray [10] have described polarization measures on distributions over one-dimensional attributes, such as income. Our measures quantify polarization of preferences – a domain that is not one-dimensional.

We provide a computational technique to calculate our metrics. The Kemeny metric is the minimal number of adjacent pairs of alternatives that have to be reversed to go from one preference to another[2]. Our metrics are determined by similar counts of pairwise exchanges. However the position in the order at which these reversals are made is relevant for us, not merely the total number of reversals. Thus, as in Kemeny's case, we have a small-dimensional statistic that characterizes the metric and it can be determined algorithmically.

2 Measuring Conflict

The space of **alternatives** is denoted $X = \{a_1, a_2, ..., a_n\}$. A **preference** π over X is identified with both an ordering of the elements of X and with a permutation of the integers $\{1, ..., n\}$. The set of all $n!$ preferences over X is denoted Π. Thus $\pi = (a_{\pi(1)} a_{\pi(2)} ... a_{\pi(n)})$ represents the preference in which a_i is preferred to a_j if and only if $\pi^{-1}(i) < \pi^{-1}(j)$. We will say in this case that a_i **precedes** a_j in the order determined by π. We will also say that if $\pi = (\pi_1 ... \pi_i ... \pi_n)$ the alternative $\pi_i = a_{\pi(i)}$ is in **position** i.

We will refer to the **natural ordering** of the alternatives as the element $e = (a_1 a_2 ... a_n) \in \Pi$ and the associated permutation is the identity $e(i) = i$ for all $i = 1, ..., n$.

Let the set of all non-empty subsets of X be denoted $\mathcal{X}$. Typically we denote a set of **feasible alternatives** by $A \in \mathcal{X}$. Given a preference π let $c_\pi(A)$ be the element in A that precedes all other elements in A. The function $c_\pi : \mathcal{X} \to X$ is the rational **choice function** generated by the preference π.

Let ν be a probability distribution over $\mathcal{X}$. We define our **measure of conflict between preferences** as

$$f(\pi, \pi'; \nu) = \nu\{A \in \mathcal{X} | c_\pi(A) \neq c_{\pi'}(A)\}$$

[1] Green-Hojman [11], Kalai-Rubinstein-Spiegler [12], Ambrus-Rosen [1], Apestigua-Ballester [2]

[2] This metric is called Kendall's τ and is a well-known measure of rank correlation.

We are going to pay special attention to the case in which the alternatives enter the feasible set in a neutral fashion. Under this assumption the probability distribution ν is invariant to any permutation of the names of the alternatives; and ν is said to be **exchangeable**. Exchangeable ν can be summarized by the distribution of the cardinality of the feasible set, or the **size distribution**, of the feasible set, $\mu = (\mu_1, ..., \mu_n)$. For every probability vector μ there is one exchangeable distribution of the feasible set and in that distribution all sets A with cardinality i are equally likely and collectively have probability μ_i.[3] [4] In treating the exchangeable case we will write μ instead of ν to specify the distribution of feasible sets under discussion.

Under neutrality, once we know the probability of disagreement between an arbitrary π and the natural ordering e the probability of disagreement between all pairs of orderings are determined. The probability of disagreement between π and e depends on ν through the size distribution μ of the feasible set A thus we can define it as

$$F(\pi; \mu) = f(\pi, e; \nu)$$

The elements of Π form a non-directed graph in which the **links** are pairs of permutations obtainable from each other by a single transposition of adjacent elements. That is (π, π') is a link if there is some $i \in \{1, ..., n\}$ such that $\pi_i = \pi'_{i-1}$, $\pi_{i-1} = \pi'_i$ and $\pi_j = \pi'_j$ for all $j \neq i-1, i$. If this relation holds we say that π differs from π' by a **transposition at position** i, the index i being the larger of the two positions at which they differ.

If π differs from e by a transposition at position i, then the family of feasible sets on which $c_\pi(A)$ differs from $c_e(A)$ is precisely those sets that contain a_{i-1} and a_i and do not contain any element a_j for $j < i-1$. For example, if π differs from e by a transposition at position n then the only feasible set on which they differ is $A = \{a_{n-1}, a_n\}$. If π differs from e by a transposition at position i, then there are 2^{n-i} sets A at which $c_\pi(A) \neq c_e(A)$ because any subset of $\{a_{i+1}, ..., a_n\}$ when combined with $\{a_{i-1}, a_i\}$ will be such a feasible set. Therefore a transposition at position i generates a collection of changes in the choice function, and we know the number of sets of each cardinality k that are affected by this transposition.

A **path** ρ from π to π' is a list of preferences $\rho_0, ..., \rho_M$ such that (ρ_{m-1}, ρ_m) is a link for all $m = 1, ..., M$, and $\rho_0 = \pi$, $\rho_M = \pi'$. Clearly there are many paths between any two permutations.

[3] deFinetti [7] showed that there is a definite structure to the family of exchangeable random sequences - countable collections of exchangeable random variables. Such sequences are essentially mixtures of independently identically distributed random variables, with different distributions. Thus, with probability one, only countable sets will arise. In the finite case there is more flexibility and the distribution of the cardinalities is unrestricted. Diaconis and Friedman [9] give error bounds in the approximation of finite exchangeable collections of random variables by subsets of exchangeable random sequences.

[4] We usually will set $\mu_1 = 0$. Choice problems with only a single feasible element are possible, but as there can be no disagreement they are trivial for our purposes. To simplify calculations we frequently assume that such problems do not arise at all.

There are many algorithms which can be applied to a permutation π that will connect it via a path to e. One of the best known of these sorting procedures is **bubble sort**.[5] In bubble sort we make a series of passes through the alternatives, beginning each time with the best (left-most) alternative. At each step in the algorithm, we compare the element in position i with the element in position $i+1$. If this pair of elements does not appear in its natural order, the elements are transposed. The algorithm continues this series of pairwise comparisons, moving from $i=1$ to $i=n-1$, at which point a pass is complete. The algorithm then begins again at $i=1$, making another pass through the ordering. The algorithm terminates when a pass is completed with no transpositions made.

In this paper, we will use a very similar algorithm: reverse bubble sort (RBS). RBS is exactly like bubble sort except that its passes move leftward from position $i=n$ until $i=2$.

While bubble sort and reverse bubble sort both convert π into e and both make only transpositions of adjacent alternatives, they generate different sequences of choice functions along the way. We choose RBS because its associated sequence has an analytically-useful property that we exploit, whereas ordinary bubble sort (and all other sorting procedures) do not share this property. The following example is an illustration.

Example 1 *Bubble Sort and Reverse Bubble Sort*

Let $n=3$ and $\pi=(a_3a_2a_1)$. As RBS sorts π, the following sequence of preferences is created. Each successive pair (ρ_{i-1},ρ_i), $i=1,2,3$ is a link because at each step only one transposition is made.

$$\begin{aligned}\rho_0 &= (a_3a_2a_1)\\ \rho_1 &= (a_3a_1a_2)\\ \rho_2 &= (a_1a_3a_2)\\ \rho_3 &= (a_1a_2a_3)\end{aligned}$$

The first link changes $c_\pi(\{a_1,a_2\})$. The second link changes $c_\pi(\{a_1,a_3\})$ and $c_\pi(\{a_1,a_2,a_3\})$. And the third link changes $c_\pi(\{a_2,a_3\})$. Note that no set A has its associated choice changed more than once.

If we had used (ordinary) bubble sort to transform π into e then we would generate the path $(a_3a_2a_1),(a_2a_3a_1),(a_2a_1a_3),(a_1a_2a_3)$. The first link would change $c_\pi(\{a_1,a_2,a_3\})$ and $c_\pi(\{a_2,a_3\})$. The second link would change $c_\pi(\{a_1,a_3\})$. And the third link would change $c_\pi(\{a_1,a_2,a_3\})$ and $c_\pi(\{a_1,a_2\})$. Note that $c_\pi(\{a_1,a_2,a_3\})$ would be changed twice along this path, first from a_3 to a_2, then from a_2 to a_1. $\square$

Example 1 illustrates a general point that is central to our analysis. We will show below that RBS never changes choice at a set A more than once. For the

[5] Bubble sort is not efficient as an algorithm. We are not concerned with computational efficiency here. As will be seen below, we are using the sorting procedure as an analytic tool.

purpose of characterizing preferences π by the choice functions they generate, we can use RBS to sort them each into e and keep track of the positions at which RBS makes a transposition. The counts of how many transpositions RBS makes as it transforms π into e will form a sufficient statistic for the family of sets at which π makes a different choice than e. These ideas are formalized as follows:

A **minimal path** from π to π' is a path $\rho_0, ..., \rho_M$ in which there is no set $A \in \mathcal{X}$ at which $c_{\rho_{m-1}}(A) \neq c_{\rho_m}(A)$ for more than one value of m. Minimal paths are useful because the set of decision problems on which π and π' disagree can be characterized by the minimal paths between them. We will now show that RBS generates a minimal path between any two preferences and we examine the families of sets where the choice function changes in more detail.

Theorem 1 *The permutations successively reached as reverse bubble sort converts π to e constitute a minimal path from π to e.*

Proof is in the Appendix.

Given π, $c_\pi(A) \neq c_e(A)$ for some family of sets A. All minimal paths from π to e change the choice at each of these sets exactly once. RBS can be used as a computational method to form a list of all sets in this family and to compute how many there are of each cardinality.

Because of our exchangeability assumption, all sets of a given cardinality have equal probability under ν. Therefore, once we know how many sets of each cardinality are changed by RBS, the probability of disagreement will be the linear combination of these counts weighted by μ.

Let us keep track of the positions at which RBS makes a transposition. In each pass, either there is or is not a transposition at position i; there can never be more than one transposition at position i in any one pass because the pass continually steps through the positions. Let x_i be the number of passes at which there is a transposition at position i. We call the vector $x = (x_2, ..., x_n)$ the RBS **signature of** π. (If it is necessary to be explicit we write $x(\pi)$ instead of x.) For example, the RBS signature in Example 1 above is $x(\pi) = (1, 2)$ because the first and second passes make transpositions at position 3 but only the first pass makes a transposition at position 2.

The RBS relationships between the six orderings of the three-alternative case can be represented geometrically. The diagram below features each ordering as a vertex of a regular hexagon. The RBS signature between each ordering and its neighbors, the two orders obtainable by a single transposition, appear around the outside of the hexagon. The element e appears in the upper left-hand corner. The arrows in the interior of the diagram are labeled with the RBS signatures for the move from the ordering e to the three non-adjacent orderings which require more than one transposition.

Insert Figure 1 *RBS Signatures for the Three-Alternative Case*

We will now show how the RBS signatures between preferences can be used to define a metric[6] on Π. This metric will be the basis for our methods in all applications. We will use it to find a central preference that best represents the preferences of a population. We also use it to measure the dispersion among preferences. Theorem 2 shows that the RBS signature between two preferences is uniquely determined at all minimal paths between these preferences. This generalizes the diagrammatic situation from Figure 1 to cases of larger n, where there can be many minimal paths. Theorem 3 then shows that the disagreement probabilities that we have discussed above define a metric on Π.

Having a metric on Π is a great analytical convenience in the computation of social choice rules and other functions on the space of distributions over Π. Specific instances of the use of metrics for this purpose are Kemeny[13], Craven[5], Klamler[15], and Barthelemy and Monjardet[4].

A converse of Theorem 3 is also valid, and we provide a proof in the appendix. Any metric on Π that is exactly additive along the minimal paths in Π can be viewed as a disagreement probability for some suitably chosen distribution over $\mathcal{X}$. The additivity property along minimal paths reflects the choice-based nature of the metric. Theorem 6 in the Appendix characterizes all such metrics, extending the axiomatization of Kemeny-Snell [14] who required additivity along all paths. Taken together, Theorems 3 and 6 show that the assumption of assent maximization is equivalent to the assumption that the procedure used for preference aggregation is a function of a metric structure on preferences.

Theorem 2 *Let $\rho_0, ...\rho_M$ be any minimal path from e to π and let y_i be the number of transpositions at position i that are made along this path. Let $y = (y_2, ..., y_n)$. Then $y = x(\pi)$.*

Proof is in the Appendix.

Theorem 3 *For any μ, the probabilities $F(\pi;\mu)$ form a semi-metric on Π. If $\mu_2 \neq 0$, then $F(\pi;\mu)$ is a metric on Π.*

Proof is in the Appendix.

We now turn to the computational use of the RBS signature in deriving the numerical values for disagreement probabilities. We will compute the number, $r(k,i)$, of sets A with cardinality k such that, if a transposition is made at position i that converts π to π' then $c_\pi(A) \neq c_{\pi'}(A)$. The transposition changes the order of two alternatives, π_{i-1} and π_i. If A has k elements in all, then $k-2$ of them must come from those $n-i$ alternatives that follow π_{i-1} and π_i in the order π. Thus

$$r(k,i) = \binom{n-i}{k-2}$$

Of course if $k-2 < 0$ then $r(k,i) = 0$.

[6] More generally, in limiting cases, the RBS signatures define only a semi-metric, although the difference will not be operationally important below. In these limiting cases we cannot distinguish between two preferences based on the choices they induce. The semi-metric induces a metric on the equivalence classes of preferences.

As all sets with cardinality k are equally likely under our assumptions, each has probability $\frac{\mu_k}{\binom{n}{k}}$. Therefore the probability w_{ik} that the feasible set is one of the sets with cardinality k where the choice is changed by the transposition at position i is

$$w_{ik} = \frac{\mu_k}{\binom{n}{k}} r(k,i)$$

Each transposition is a measure of disagreement between π and e. The algorithm changes the choice at a family of feasible sets from $c_\pi(A)$ to $c_e(A)$ which is precisely the family at which π and e disagree, and it never makes a change where $c_\pi(A) = c_e(A)$. The minimality of the path assures that the total disagreement is partioned among the links with no double counting. Thus, the probability that the feasible set is such that π and e disagree can be computed by adding the disagreement probabilities at every step of the RBS algorithm. Adding the probabilities as computed above and weighting each of them by the number of transpositions required at each position we have:

Theorem 4 *If π has RBS signature $x(\pi)$ and ν is exchangeable with size distribution μ then the probability of disagreement between π and e is given by*

$$F(\pi;\mu) = \sum_{i=2}^{n} \sum_{k=2}^{n} w_{ik} x_i = \sum_{i=2}^{n} x_i(\pi) \sum_{k=2}^{n} \frac{\mu_k}{\binom{n}{k}} \binom{n-i}{k-2} \qquad (1)$$

This Theorem follows from the above discussion of counts of sets by cardinality.

Formula (1) shows how π and the size distribution μ enter into the probability of disagreement. The effect of the preference π is completely summarized by its RBS signature $x(\pi)$. This disagreement probability is a linear function of the signature, with coefficients dependent on the size distribution of the feasible sets.

We now examine some special cases of exchangeable distributions and see what this formula tells us.

The case in which it is sure that the feasible set will contain exactly two alternatives is the problem studied by Kemeny [13]. In this case we define

$$\mu^K = (0,1,0,...,0)$$

. We know from the Kemeny [13] and Young and Levenglick [20] papers that in this case F is Kendall's tau: $w_{ik} = 1$ for all k and $\sum_{i=2}^{n} x_i$ is the total number of pairwise exchanges needed to carry π into e.

In the case in which all $A \in \mathcal{X}$ with two or more elements are equally likely, a transposition at position i affects the optimum if and only if both π_{i-1} and π_i are in A and none of π_k are in A for $k < i-1$. The probability of this event is $\frac{1}{2^i}$. Under this distributional assumption we define

$$\mu^E = (\sum_{j=2}^{n} \binom{n}{j})^{-1} (0, \tbinom{n}{2}, ... \tbinom{n}{n})$$

The scalar factor adjusts for the fact that we put zero probability on sets of order zero and one.

Finally, in the case where it is certain that all alternatives will be available, so that $A = X$ with probability one, we define

$$\mu^X = (0, ..., 0, 1)$$

Two preferences π and π' disagree if and only if $\pi_1 \neq \pi'_1$. Thus $w_{2n} = 1$ and $w_{ik} = 0$ for $i > 1$ or $k < n$. In this case F assigns zero distance between all preferences that agree with e on the top element, and a unit distance for all those that disagree. It is in cases such as this that F is a semi-metric instead of a metric, because only top elements matter to choice.[7]

3 Assent-maximizing Social Welfare Functions

Having identified conflict with the mathematical expectation of disagreement between preferences, we now look for social orderings that minimize expected conflict. Specifically, we look for a social ordering π^* that maximizes the probability that its choice from a random given subset agrees with the choice that would be made by a randomly-selected member of the population.

We discuss the relationship of the social ordering that is produced by our measures with that produced by the Kemeny method. We will emphasize the contrast between μ^K and the special case of equally-likely feasible sets μ^E. Similar examples can be constructed for many other size distributions. The important point is that larger potential feasible sets will generate a different pattern of agreement and disagreement than if only pairs are possible.

Imagine a population of individuals, each of whom holds a particular preference over X. Following the classical social choice approach, we invoke anonymity in our treatment of these individuals. Therefore, we can represent this population as a distribution over Π. Denote a particular **population** by λ and let Λ be **set of all populations**.

Given a population λ, we can describe the level of expected conflict between a randomly-drawn member of the population and any potential candidate for a social ordering. Let the **frequency of expected conflict between a candidate social ordering π' and individual preferences in the population λ** be defined by

$$q(\pi', \lambda, \nu) = \sum_{\pi} f(\pi, \pi'; \nu)\lambda(\pi) \tag{2}$$

This expression is the key argument in our social aggregation procedure. The **assent-maximizing social welfare function** is defined as the social ordering

[7] Craven (1996) examines the social choice rule associated several metrics on the space of orderings different from μ^K following the Kemeny's approach in other respects. One of his metrics is μ^X, corresponding to the case in which it is certain that all alternatives are available. The other metrics are among those described in Diaconis (1988). They do not lead to an interpretation as the probablity of disagreement and are thus not "choice-based" in our terminology.

that minimizes this expected conflict

$$\pi^*(\lambda, \nu) = \arg\min_{\pi'} q(\pi', \lambda, \nu) \tag{3}$$

If ν is exchangeable and we want to emphasize the dependence of social choice on the size distribution μ of the set of alternatives, we will write the assent-maximizing social welfare function as a function of μ, with the slight abuse of notation $\pi^*(\lambda, \mu)$.

When (2) is evaluated at the assent-maximizing social welfare function we have the **measure of conflict with assent-maximization**

$$Q(\lambda, \nu) = q(\pi^*(\lambda, \nu), \lambda, \nu) = \sum_{\pi} f(\pi, \pi^*; \nu)\lambda(\pi) \tag{4}$$

The measure Q corresponds to the "goodness of fit" of the social preference to the population of individual preferences.

Given a population λ, we can also define the **measure of internal conflict**

$$\bar{Q}(\lambda, \nu) = \sum_{\pi} \sum_{\pi'} \lambda(\pi)\lambda(\pi') f(\pi, \pi'; \nu) \tag{5}$$

The measure $\bar{Q}$ does not depend on a social ordering. It is a direct measure of the diversity of preferences within the population.

Example 2 *Maximizing Assent to Plurality-Induced Choices*

This is our basic example where we show that the Kemeny method, which maximizes assent under the assumption that only pairs are possible, can lead to substantially less average assent then the true assent-maximizing social welfare function, in cases where larger feasible sets are likely. We take the case of equally-likely feasible sets for concreteness, with $n = 3$ and $\mu = \mu^E = (0, \frac{3}{4}, \frac{1}{4})$.

The population λ is concentrated on three of the six orderings in Π according to the distribution: $\lambda((a_1 a_2 a_3)) = .49$, $\lambda((a_3 a_2 a_1)) = .48$, $\lambda((a_2 a_3 a_1)) = .03$. This distribution exhibits the tension between selecting the ordering consistent with the transitive majority relation (Condorcet Consistency) and selecting the Plurality winner from the three-alternative set $A = X$. Using Plurality rule on every subset, the choice function would be

$$\begin{aligned} c(\{a_1, a_2\}) &= a_2 \\ c(\{a_1, a_3\}) &= a_3 \\ c(\{a_2, a_3\}) &= a_2 \\ c(\{a_1, a_2, a_3)\} &= a_1 \end{aligned}$$

Condorcet Consistency requires that the social ordering be $(a_2 a_3 a_1)$ as the majority choices from the pairs are transitive. But this ordering does not describe the result of plurality rule when all three alternatives are available, as a_1 not a_2 is the plurality choice from the order three subset. This illustrates an important point: populations that satisfy Condorcet Consistency will not necessarily generate plurality choice functions that are consistent with an ordering. Thus, if the social choice procedure must produce an ordering, it will have to make tradeoffs between matching plurality choices on subsets of various sizes. The Kemeny method selects the ordering $\pi = (a_2 a_3 a_1)$, which implies a choice of $c(\{a_1, a_2, a_3\}) = a_2$, supported by just 3% of the population. If $A = \{a_1, a_2, a_3\}$ never arises, then this choice from the triple does not create any expected dissent. If $A = \{a_1, a_2, a_3\}$ may arise, however, expected assent may be maximized by an ordering that has a_1 as its top element instead of a_2. When all subsets with two or more elements are equally likely, we have $\pi^*(\lambda, \mu^E) = (a_1 a_2 a_3)$, implying a choice of a_1 from $A = \{a_1, a_2, a_3\}$ that is supported by 49% of the population. This gain of 46% in expected support for the social choice at $A = \{a_1, a_2, a_3\}$ more than outweighs the decrease of 1% or 2% in support that occurs when $\{a_1, a_2\}$ or $\{a_1, a_3\}$ are the feasible sets. Assent-maximization with μ^E sacrifices matching the majority preference on two of the three pairs problems in order to better match the population preferences in the case where all three alternatives are available. □

Example 2 emphasizes the tension between assent-maximization and the ordering requirement. In these cases we need to consult the distribution ν when forming a social ordering. On the other hand, when the plurality choice from each A is explainable by an ordering, that is, when there exists $\hat{\pi}$ such that $c_{\hat{\pi}}(A)$ coincides with the plurality winner at every $A \in \mathcal{X}$, then this tension does not exist and, as the following Theorem states, the social ordering should be independent of ν - in fact it should be the same as the ordering $\hat{\pi}$.

Theorem 5 *If $\lambda \in \Lambda$ is a population of preferences such that the assent-maximizing social welfare function $\pi^*(\lambda, \nu)$ is independent of ν then plurality choice at each $A \in \mathcal{X}$ coincides with $\pi^*(\lambda, \nu)$. Conversely, if plurality choice at $A \in \mathcal{X}$ is a rational choice function, then this choice function is the assent-maximizing social welfare function for every distribution ν.*

Proof is in the Appendix.

The condition that the plurality choice function be derivable from an ordering is very strong. The simplest type of population which generates a rational plurality choice function contains a majority held preference, where $\lambda_i > .5$ for some π_i. It is obvious in this case that the choice funtion induced by π_i maximizes assent at every subset $A \in \mathcal{X}$, implying that π_i is assent-maximizing for any distribution ν. Populations with a majority-held preference are a strict subset of the populations which satisfy the premises of Theorem 5. Note that if we restrict attention to only exchangeable distributions, this result would not hold. It is possible that the assent-maximizing ordering is invariant over the smaller space of exchangeable distributions and yet the plurality choice function is not rational.[8].

[8] An example of this situation is the following. Take a population λ at which plurality rule

We now concentrate on the typical case in which there is a tension between assent-maximization and the ordering requirement. Through a series of four examples we will demonstrate some features of assent-maximizing social welfare functions. We will show that assent-maximization may select something other than a Condorcet winner, even when one exists. We show that the assent-maximizing ordering may lie outside the support of the preferences in the population. We show that the method of compromise inherent in assent-maximization is different from both the compromises made by positional methods such as scoring rules and the compromises made by methods that use only the numerical vote tallies from pairwise contests. Finally we explore the relationship between the dispersion around the social choice and the measure of average disagreement across pairs in the population, showing that these variance like measures can change their order depending on the probabilities ν.

Example 3 *Assent-maximization may not rank a Condorcet winner first*

This is a well-known four-alternative problem, that of chosing the capital city of Tennessee (see Young [19], Moulin[16]), which we adapt for the present discussion by allowing the feasible sets to be larger than pairs. For concreteness, we assume that all subsets are equally likely.

The four alternatives are the four largest cities in Tennessee:

Memphis $= a_1$, Nashville $= a_2$, Chattanooga $= a_3$, Knoxville $= a_4$

The distribution of preferences λ is: $\lambda((a_1a_2a_3a_4)) = .42$, $\lambda((a_2a_3a_4a_1)) = .26$, $\lambda((a_3a_4a_2a_1)) = .15$, $\lambda((a_4a_3a_2a_1)) = .17$

We have $\pi^*(\lambda, \mu^E) = (a_1a_2a_3a_4)$ and $\pi^*(\lambda, \mu^K) = (a_2a_3a_4a_1)$. The difference between our method and the Kemeny method is based on a rather narrow difference in average assents to the social choice: $q((a_1a_2a_3a_4), \lambda, \nu^E) = .5282$, but $q((a_2a_3a_4a_1), \lambda, \nu^E) = .5173$. This is an admittedly small difference in average assent. Nevertheless, the induced choice functions are quite different

would generate cyclic choice over the pairs in a three-alternative problem:

$\lambda(a_1a_2a_3) = \frac{1}{3} - \varepsilon$, $\lambda(a_2a_3a_1) = \frac{1}{3} + \varepsilon$, $\lambda(a_3a_1a_2) = \frac{1}{3}$, where ε is small and positive. The claim is that $\pi^* = (a_2a_3a_1)$ is the assent-maximizing order for all exchangeable distributions. Consider the following table of assents to choices from the pair subsets for each potential social ordering.

A	π_1	π_2	π_3	π_4	π_5	π_6
$\{a_1, a_2\}$	$\frac{2}{3} - \varepsilon$	$\frac{2}{3} - \varepsilon$	$\frac{1}{3} + \varepsilon$	$\frac{1}{3} + \varepsilon$	$\frac{2}{3} - \varepsilon$	$\frac{1}{3} + \varepsilon$
$\{a_2, a_3\}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\{a_1, a_3\}$	$\frac{1}{3} - \varepsilon$	$\frac{1}{3} - \varepsilon$	$\frac{2}{3} + \varepsilon$	$\frac{1}{3} - \varepsilon$	$\frac{2}{3} + \varepsilon$	$\frac{2}{3} + \varepsilon$
Sum	$\frac{5}{3} - 2\varepsilon$	$\frac{4}{3} - \varepsilon$	$\frac{5}{3} + 2\varepsilon$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{4}{3} + 2\varepsilon$

For an exchangeable distribution, the assent to each pair is weighted equally in the maximization. Thus, the final row of sums captures the relevant information about assent to the pairs. We see π_3 strictly dominates the other candidates on pairs. The total expected assent for any candidate for any exchangeable distribution will be a weighted sum of the pairs assent and the triple assent. Since a_2 is the assent-maximizing choice from the order three subset, π_3 weakly dominates the other candidates on the triple. Therefore π_3 is assent-maximizing at all exchangeable distributions. But, the plurality choice function of this population is not rational because of the cyclic relationship on the pairs.

By Theorem 5, there are non-exchangeable distributions at which π_3 would not be assent-maximizing. For example, a non-exchangeable distribution that put most of the weight on $\{a_1, a_2\}$ would not produce π_3 as a social ordering.

because $(a_1a_2a_3a_4)$ selects Memphis whenever it is available, whereas $(a_2a_3a_4a_1)$ never puts the capital in Memphis at any feasible set.

Notice that Nashville (a_2) is a Condorcet winner, beating the other three alternatives in the three pairwise contests where it is one of the options. For that reason the Kemeny method must rank a_2 first. On the other hand, in every available set with three or more alternatives in which a_1 is one of the options, it is chosen by the plurality of the population. Therefore as long as the probability is high enough that the actual available set will have three or more options, a_1 should be chosen from these sets. In these cases, therefore, a_1 should be first in the social ordering. $\square$

Example 4 *Compromising on a Second Best*

In both Example 2 and Example 3 the assent-maximizing π^* was one of the orderings in the support of λ. This is not always the case as the following four alternative example shows:

Let $\lambda((a_3a_1a_2a_4)) = .33$, $\lambda((a_2a_1a_3a_4)) = .34$, $\lambda(a_4a_1a_2a_3) = .33$. For a variety of metrics, including the equally-likely size distribution μ^E, $\pi^*(\lambda, \mu) = (a_1a_2a_3a_4)$. This social ordering is somewhat of a compromise of the three orderings in the population; it places the alternative that everyone in the population ranks second, a_1, at the top of the social ordering. Though this means that no one in the population will agree with the choice from the order four subset, this social ordering will perform better on average on subsets of order two and three than an ordering chosen to match the first-place choice of one segment of the population. The conflict-minimization method trades off support on the order four subset in order to garner greater support on the more likely order two and three subsets. Note, however, that if the probability of facing the order four subset was sufficiently high, the social ordering would change to $\pi = (a_2a_1a_3a_4)$. $\square$

The social choice rules based on assent maximization differ from the two main classes of methods that have been applied and characterized in the literature. These methods are the scoring rules, which produce an ordering by combining rank information from the individual preferences, and pairwise methods that produce an ordering as a function of the majority tournament. Theorem 5 demonstrates assent maximization is not equivalent to any scoring rule. The reason is that scoring rules will not always produce a social ordering coincident with that of majority of the population. A minority with very different preferences will be able to influence the relative scores.[9] We can also see that assent maximization is not a member of the class of social welfare functions based only

[9] A population at which a strict majority would not dominate the social ordering under a scoring rule is $\lambda((a_1a_2a_3a_4)) = .55$, $\lambda((a_2a_4a_3a_1)) = .25$, $\lambda((a_2a_3a_4a_1)) = .2$. For all ν our method reproduces the majority's preference $\pi^* = (a_1a_2a_3a_4)$ but Borda count produces $(a_2a_1a_3a_4)$ because the minority preferences give a_2 an advantage over their last-choice, a_1.

on pairwise comparisons and satisfying neutrality. There can well be a tension between the way that pairwise methods break a Condorcet cycle, at its weakest link, and the plurality choice from a larger feasible set. The following example shows that the assent maximizing ordering can differ from the unique recommendations of both of these classes of methods, and thus that the social welfare function it defines does not lie in either class.

Example 5 *Positional and Pairwise Methods and Assent-Maximization*

In this example we show that for a particular population, with preferences over just three alternatives, positional methods, pairwise methods, and our method for a range of μ can generate three distinct social orderings.

Consider the population λ defined by

$$\begin{aligned} \lambda((a_1a_2a_3)) &= .10 \\ \lambda((a_1a_3a_2)) &= .30 \\ \lambda((a_3a_2a_1)) &= .25 \\ \lambda((a_2a_1a_3)) &= .35 \end{aligned}$$

Saari[17] has shown that all positional methods will select $(a_1a_2a_3)$[10], while Kemeny's method (and all pairwise-based methods) will select $(a_2a_1a_3)$. Our rule, for sufficiently large μ_3, will yield $(a_1a_3a_2)$.[11] This range of μ places a probability on the three-alternative choice problem greater than the .25 assigned by μ^E. As a result, there is a range of μ under which alternative a_1 is in first-place, as it will be supported as the choice from the triple by .40 of the voters. Alternative a_2 from the triple would be supported by just .35 of the population, and a choice of alternative a_3 is supported by just .25. Thus, with respect to the first-place choice, our method looks similar to the positional methods. Now consider the ordering of the final two alternatives. Since .55 of the population of voters prefer a_3 to a_2 our method selects the ordering (a_3a_2), respecting the will of the majority on this pair. Positional methods select the ordering $(a_1a_2a_3)$. Thus no positional method or pairwise based method will agree with our methods over the entire range of μ. $\square$

Example 6 *Conflict and Compromise*

Let λ and λ' be two populations such that for a given μ , λ is less internally conflicted than λ': $\bar{Q}(\lambda, \nu) < \bar{Q}(\lambda', \nu)$ Will there be more assent to the assent-maximizing social ordering at λ than there will be at λ'? In other words, is it easier to find a good compromise when the population is less conflicted?

[10] Saari's result can be seen as follows: Consider any scoring rule, standardized to the vector $(1, s, 0)$ where $s \in [0, 1]$. Then, we have the following scores for a, b, c, respectively: $8 + 7s, 7 + 7s, 5 + 6s$.

[11] The value of μ_3 needed for this result is well in excess of .25 which is μ_3^E.

A simple example shows that this will not generally be the case. We will fix $\mu = \mu^E$ and let λ and λ' be given by:

$$\begin{aligned}
\lambda((a_1a_2a_3)) &= .5 \\
\lambda((a_2a_1a_3)) &= .5 \\
\\
\lambda'((a_1a_2a_3)) &= .5 \\
\lambda'((a_1a_3a_2)) &= .25 \\
\lambda'((a_2a_1a_3)) &= .25
\end{aligned}$$

Insert Figure 2 *Conflict and Compromise*

It is straightforward to show that λ' is more internally conflicted than λ.

$$\begin{aligned}
\bar{Q}(\lambda, \nu) &= .0625 \\
\bar{Q}(\lambda', \nu) &= .0703
\end{aligned}$$

But

$$\begin{aligned}
Q(\lambda, \nu) &= .125 \text{ with } \pi^*(\lambda) = \{(a_1a_2a_3), (a_2a_1a_3)\} \\
Q(\lambda', \nu) &= .0938 \text{ with } \pi^*(\lambda') = \{(a_1a_2a_3)\}
\end{aligned}$$

The reason for this difference in conflict measures can be traced to the fact that the space of preferences is discrete. The measure of conflict with the assent-maximizing ordering is like a second moment taken around a point in the space which is not a true mean – one that is chosen in asymmetric situations because it is an approximate compromise and lies within the space from which compromises must be selected. The measure of internal conflict, on the other hand, deals directly with pairs of preferences chosen from the population and does not require any such point of reference. □

4 Explanations of Irrationality and the Measurement of Welfare

This section relates to the work of Green and Hojman [11] on the measurement of welfare when the decision-maker is irrational. Green and Hojman analyze a decision-maker who displays an arbitrary choice function $c : \mathcal{X} \rightarrow X$ that is, in general, inconsistent with the maximization of any single preference relation. Green and Hojman assume that the decision-maker has multiple objectives, holding each with a possibly different strength. Thus there is a distribution of preferences $\lambda \in \Lambda$ that is welfare relevant. As in the case of ordinary welfare economics, we wish to infer λ from the observation of c. However, we cannot identify from the data the way in which the objectives were aggregated to determine the choice function. Let $\mathcal{C}$ be the set of all choice functions and let $v : \Lambda \times \mathcal{X} \to \mathcal{C}$ be an **aggregation rule** that describes how diverse preferences are combined to make a choice from any feasible set. The aggregation rule is a correspondence, allowing multiple choices at some boundary situations. Green and Hojman then define an **explanation for a choice function** c as distribution over preferences and an aggregation rule (λ, v) such that for all $A \in \mathcal{X}$, $c(A) \in v(\lambda, A)$. They assume that λ is welfare-relevant and v is not. Therefore, for welfare purposes we are interested in identifying the set of all $\lambda \in \Lambda$ that are part of some explanation of c. This set of preference distributions depends on the family of aggregators that are allowed. The richer the class of aggregators the more populations of preferences can be consistent with c.[12]

Even if the allowable aggregators are a tightly specified family there is still considerable indeterminacy in the explanatory λ in this theory. The idea explored in this section is that one way to select a good candidate for welfare analysis is to choose the λ with the least internal conflict, among those that can explain c. Our measure of internal conflict will be generated by the metrics developed above.[13]

Why is the selection of a conflict-minimizing explanation a good choice for a welfare measurement method? Two principal reasons can be given. The first is Occam's Razor. The simplest explanation is to be preferred; conflict minimization is one way of defining simplicity.

The second reason is by analogy to what is done in ordinary welfare economics with a rational choice function. When the choice function c is rational – that is, there exists, π such that $c(A) = c_\pi(A)$ for all $A \in \mathcal{X}$, economists say that c is "explained" by the "revealed preference" π. In this favorable circumstance, π is used for all welfare inferences. However, if one admits that the decision-maker may simultaneously hold multiple objectives then the selection of the λ that is a unit mass at π becomes only one possibile explanation among many. Any other λ' that produces the same choice function c when aggregated would also qualify, even though λ' may give significant weight to "minority preferences" that conflict with λ at the observed choice at some A. By concentrating on the λ which is a point mass at π, ordinary welfare economics is selecting the conflict-minimizing explanation. In this section, therefore we examine the use of conflict minimization as a selection criterion in the general irrational case.

[12] Green and Hojman use the family of all scoring rules and the richer family of all monotonic rules, obtaining similar results in the two cases. Truly perverse behavior, in which choice responds non-monotonically to changes in the preferences that are held, will make welfare inferences based on choice behavior essentially impossible.

[13] Other authors use selection criteria different from conflict minimization. For example, de Clippel-Eliaz [6] use explanations that have at most two preferences in their support, and Ambrus-Rozen [1] use explanations with the minimal cardinality of preferences and aggregators that can incorporate cardinal information. It would be interesting to determine the precise relationship between such cardinality-minimizing criteria and our conflict-minimizing criterion.

We now give two examples showing how this approach leads to a much more tightly-specified selection for welfare purposes. Both examples will deal with well-known three-alternative cases.

Example 7 *Explaining Second-Place Choice*

In this example the choice function is consistent with rational choice on problems with two alternatives, but when all three alternatives are available the choice is the one that seems to be ranked second based on the preference revealed in pairwise problems. This choice function is frequently used to illustrate the compromise effect in psychology.

$$\begin{aligned} c(\{a_1, a_2, a_3\}) &= a_2 \\ c(\{a_1, a_2\}) &= a_1 \\ c(\{a_1, a_3\}) &= a_1 \\ c(\{a_2, a_3\}) &= a_2 \end{aligned}$$

Insert Figure 3 *Explaining Second-Place Choice*

Let the set of allowable aggregation rules be the family of scoring rules. Consider the two populations $\lambda((a_1a_2a_3)) = \frac{1}{2}$, $\lambda((a_2a_1a_3)) = \frac{1}{2}$ and $\lambda'((a_1a_2a_3)) = \frac{1}{2}$, $\lambda'((a_2a_3a_1)) = \frac{1}{2}$. Both will generate second-place choice as long as the proper selection is made from multi-valued $v(\lambda, A)$. For example, if v is Borda count then $v(\lambda, \{a_1, a_2\}) = \{a_1, a_2\}$, $v(\lambda, \{a_1, a_2, a_3\}) = \{a_1, a_2\}$, $v(\lambda', \{a_1, a_2\}) = \{a_1\}$, $v(\lambda', \{a_1, a_2, a_3) = \{a_2\}$ and therefore both (λ, v) and (λ', v) are explanations of c.

When we compare these populations using the criterion of least internal conflict we see that for all metrics (except the limiting semi-metric case of $\mu = (0, 0, 1)$), λ is less internally conflicted than λ'. Using the criterion of minimal internal conflict to select among explanations is telling us to maintain the agreement between members of the explanatory population on the pair $\{a_1, a_3\}$ whenever it is possible to do so. Despite the fact that we have no evidence of an internal conflict, as a_3 is never chosen by c in any circumstance, we also have no direct evidence that all preferences in the population do agree on this set. Using the explanation (λ', v) would introduce the prospect of a disagreement over $\{a_1, a_3\}$. The criterion of minimal internal conflict tells us not to invoke such explanations unless it is necessary to do so.

Example 8 *Explaining Cyclic Choice*

The second example is the standard cyclic choice pattern with three alternatives:

$$\begin{aligned} c(\{a_1, a_2, a_3\}) &= a_1 \\ c(\{a_1, a_2\}) &= a_1 \\ c(\{a_1, a_3\}) &= a_3 \\ c(\{a_2, a_3\}) &= a_2 \end{aligned}$$

Insert Figure 4 *Explaining Cyclic Choice*

As in the case of the example above there are many populations that are part of explanations of c. Two candidates that can produce cyclic choice at all scoring rules are:

$$\begin{aligned} \lambda((a_1a_2a_3)) &= \frac{1}{2},\ \lambda((a_3a_2a_1)) = \frac{1}{2} \\ \lambda'((a_1a_2a_3)) &= \frac{1}{3}, \lambda'((a_2a_3a_1)) = \frac{1}{3}, \lambda'((a_3a_1a_2)) = \frac{1}{3} \end{aligned}$$

It is straightforward to compute that $\bar{Q}(\lambda, \nu^K) > \bar{Q}(\lambda', \nu^K)$ but that $\bar{Q}(\lambda, \nu^E) = \bar{Q}(\lambda', \nu^E)$. Moreover, if ν is such that the probability that $A = X$ is higher than $\frac{1}{4}$ then $\bar{Q}(\lambda, \nu) < \bar{Q}(\lambda', \nu)$. Therefore, unlike the case of second-place choice, the less conflicted explanation for cyclic choice depends on the size distribution of the problems that are faced. It is less clear which explanation of cyclic choice is simpler, if by simplicity we mean a smaller measure of internal conflict.

5 Polarization and Conflicting Preferences

In this section we discuss the application of our metrics to the measurement of polarization in the spirit of Esteban and Ray [10] (ER). ER have pioneered the axiomatic study of polarization and have distinguished it from the measurement of diversity, inequality or heterogeneity in a population. The polarization of a population depends on the sub-populations that comprise it.[14] As ER describe it there are two questions to be asked. The first is how united the sub-populations are likely to be. The second is how distinct they are from each other. They call these variables **identification** and **alienation**.

Identification is more likely if a group is large and if it is composed of similar individuals. It would then be easier for people in this group to find and

[14] ER describe a situation where there are two exogenous sub-populations between which a conflict may arise. Their index, however, treats individuals as if they have no group identity except for their income and tries to predict conflict based on measured polarization of the income distribution.

recognize others, and to cooperate with them in a conflict. A small but highly unified sub-population is unlikely to engage in conflict because they are too small a proportion of the whole to succeed. A large, highly heterogeneous sub-population will find conflict difficult because it may not cohere. Coordination may be difficult and it may even break apart due to internal frictions.

Alienation measures the difference between those in a group and those not in the group. In the case of income differences, or that of any other one-dimensional attribute as discussed by ER, the natural measure of alienation is the absolute value of the difference in this attribute.

In this section we do three things. First we show how to extend the polarization indices of ER to the political sphere by applying our metrics on ordinal preferences. Second we show how our metrics can be useful for modifying the measurement of identification and alienation, as propsed by ER. Third we show how the choice of a particular metric within the class we have characterized above may affect the comparison of polarization across two populations. Using a numerical example we explain the choice-theoretic basis for this variation across metrics.

ER consider a discrete distribution of incomes y where y_i represents the income of group i and γ_i is the proportion of this group in the whole population. The measure of polarization that ER derive axiomatically is

$$\sum_i \sum_j \gamma_i^{\kappa} \gamma_j |y_i - y_j| \text{ where } \kappa \in (1, 2.6]$$

The term $\gamma_i \gamma_j |y_i - y_j|$ captures alienation by computing the expected absolute difference between randomly selected members of the population. To the extent that $\kappa > 1$, the additional multiplicative term of $\gamma_i^{\kappa-1}$ captures identification recognizing that larger income groups will find it disproportionately easier to join forces. ER provide axioms to derive this multiplicative form for the polarization index, based on these measures of identification and alienation.

Our metrics enable us to modify the ER specification of identification to take into account both group size and similarity of group members and to capture alienation as the average difference between ordinal preferences. Through these modifications we try to capture the spirit of the E-R model.

We consider two exogenously-defined groups in the population denoted A and B and let $\lambda^A(\pi)$ and $\lambda^B(\pi)$ be the portions of the population with preference π in each of the two groups. Then one way of creating an index of identification based on our metrics could be

$$\begin{aligned} I^A(\lambda^A) &= (\sum_{\pi} \lambda^A(\pi))^{\kappa} [\sum_{\pi} \sum_{\pi'} \lambda^A(\pi) \lambda^A(\pi')(1 - f(\pi, \pi'; \nu)] \\ I^B(\lambda^B) &= (\sum_{\pi} \lambda^B(\pi))^{\kappa} [\sum_{\pi} \sum_{\pi'} \lambda^B(\pi) \lambda^B(\pi')(1 - f(\pi, \pi'; \nu)] \end{aligned}$$

This identification index depends on the size of the group, as determined by the parameter κ as in the ER theory and homogeneity within the group. We measure homogeneity by the probability of agreeing on choice from a feasible set: $1 - f(\pi, \pi'; \nu)$. The average agreement probability in the group is a measure of its cohesiveness. By multiplying the non-linear index of group size by this index of cohesiveness we obtain a plausible identification index.

Similarly, an index of alienation can be obtained from across-group differences. The idea is that each person considers how different a randomly-selected member of the other group would be in comparison to a randomly-selected member of his own group. Let $\eta^A = \sum_\pi \lambda^A(\pi)$ be the fraction of the whole population that is in group A, so that $\lambda^A(\pi)/\eta^A$ is the conditional distribution of preferences within group A. Then the alienation felt by group A relative to group B is

$$a^A(\lambda^A, \lambda^B) = \sum_\pi \lambda^A(\pi) \sum_{\pi'} |\frac{\lambda^A(\pi')}{\eta^A} - \frac{\lambda^B(\pi')}{\eta^B}| f(\pi, \pi'; \nu)$$

And symmetrically,

$$a^B(\lambda^B, \lambda^A) = \sum_\pi \lambda^B(\pi) \sum_{\pi'} |\frac{\lambda^A(\pi')}{\eta^A} - \frac{\lambda^B(\pi')}{\eta^B}| f(\pi, \pi'; \nu)$$

If the two sub-populations have the same distribution of preferences then there is no alienation because the two conditional distributions are identical. Similarly, if disagreement probabilities are low then there is little alienation, even if the two groups are large and distinct.

Based on these indices of identification and alienation we can define a measure of polarization as

$$P(\lambda^A, \lambda^B) = (I^A(\lambda^A) \cdot a^A(\lambda^A, \lambda^B)) \cdot (I^B(\lambda^B) \cdot a^B(\lambda^B, \lambda^A))$$

The reason for taking a multiplicative structure for the polarization measure is that polarization requires both alienation and identification. If the two groups have no alienation because they have the same conditional distributions of preferences then there should be no polarization regardless of how well-identified they are. Likewise, if either of the two groups has low identification, either because of small size or lack of cohesion, then there should be little polarization because conflict is unlikely among poorly identified groups. Conflict is highest when the two groups are distinct in preference and divide the population into two large sub-groups of similar individuals.[15]

Example 9 *Dependence of Polarization on the Metric*

In this example we give two populations and compute the polarization index above using different metrics from our family. We show how their relative levels of polarization vary with the choice of the metric, and we give a choice-theoretic interpretation for this outcome.

[15]Other polarization indices can have different, related functional forms. For example flexible forms that allow for tradeoffs between identification and alienation at a given level of polarization should be explored.

The first population contains two subgroups, which we will denote A and B. Subgroup A has a mass of .5, all of which is concentrated on the preference $(a_1a_2a_3a_4)$. Subgroup B consists of a mass of .25 of the population with the preference $(a_2a_1a_4a_3)$ and a mass of .25 with the preference $(a_4a_1a_3a_2)$.

The second population also contains two subgroups, denoted A' and B'. As in population 1, subgroup A' has a mass of .5, all of which is concentrated on the preference $(a_1a_2a_3a_4)$. In this population, however, subgroup B' consists of a mass of .25 with the preference $(a_3a_1a_2a_4)$ and a mass of .25 with the preference $(a_4a_2a_1a_3)$.

Let us take a detailed look at the choices made by these four subgroups at each of the eleven subsets of X with two or more elements. In population 1, the two halves of subgroup B, $(a_2a_1a_4a_3)$ and $(a_4a_1a_3a_2)$, disagree with subgroup A on 5 and 8 of the 11 subsets of X, respectively. In population 2, disagreement is slighly more prevalent, as the two halves of subgroup B', $(a_3a_1a_2a_4)$ and $(a_4a_2a_1a_3)$, disagree with subgroup A' at 6 and 9 out of the 11 subsets, respectively. However, when we restrict attention to pairs only, we see that both subgroup B and subgroup B' have the same pattern of disagreements – each has a half that disagrees with A (or A') on 2 out of 6 pairs and a half that disagrees with A (or A') on 4 out of 6 pairs.

Computing the polarization indices using the equally likely metric, we see that (A', B') is more polarized than (A, B), as

$$\frac{P(\lambda^A, \lambda^B)}{P(\lambda^{A'}, \lambda^{B'})} = \frac{13^3}{15^2 \cdot 11} < 1$$

However, if the Kemeny metric were used then we have

$$\frac{P(\lambda^A, \lambda^B)}{P(\lambda^{A'}, \lambda^{B'})} = \frac{4}{3} > 1$$

This is because under the Kemeny metric alienation is the same across (A, B) and (A', B'). The difference between these populations lies in the identification of the subgroups. Subgroup B' is less identified because its two parts have completely opposed preferences, whereas in Subgroup B the two preferences are very different, but not maximally so. Since polarization increases in more identified societies, holding constant the level of alienation, we have the indicated result.

Based on these two calculations one might think that as the metric puts more weight on larger sets the relative polarization of (A, B) will increase relative to that of (A', B'). This conjecture is false, however, because if we use the metric derived from $\mu = (0, 0, 1)$, where only the first-choice of each group matters because all alternatives are always available, we see that both alienation and identification are the same in the two populations. Thus, polarization behaves non-monotonically with respect to the distribution of probabilities over problems of different sizes. These types of surprising non-monotonicities highlight the subtleties of measuring polarization in multi-dimensional domains. The dependence of the measure on the size distribution of the feasible set is an important theoretical observation with the potential to inform empirical work on polarization.□

Example 10 *Polarization versus Internal Conflict*

The following example illustrates the difference between polarization and the measure of internal conflict of a population. Let λ and λ' be two populations defined as follows:

$$\begin{aligned}\lambda((a_1a_2a_3)) &= \frac{1}{2},\, \lambda((a_3a_3a_1)) = \frac{1}{2} \\ \lambda'(\pi) &= \frac{1}{6} \text{ for all } \pi\end{aligned}$$

Insert Figure 5 *Polarization versus Internal Conflict*

The sub-populations we consider are $A = \{(a_1a_2a_3), (a_1a_3a_2), (a_2a_1a_3)\}$ and $B = \{(a_2a_3a_1), (a_3a_1a_2), (a_3a_2a_1)\}$. In this example λ is concentrated on one representative preference from each of the two sub-populations A and B, whereas λ' has all preferences equally represented. Population λ is more polarized than λ' because the diversity within λ' reduces identification.

A straightforward calculation reveals that λ is less internally conflicted than λ' according to our measure of internal conflict $\bar{Q}$. The measure $\bar{Q}$ averages our metric over all pairs of preferences, without reference to the sub-populations from which these preferences are drawn. The distinction emphasized by ER is upheld by our measures as well. Internal conflict captures the overall diversity in a population whereas polarization measures potential conflict across sub-populations.

6 Conclusion

We have introduced a new family of metrics on the space of ordinal preferences, based on the choice-theoretic implications of these preferences when they face randomly selected decision problems. For the case in which the decision problems are selected using exchangeable distributions we give an algorithmic method to find the distance between two preferences

Our principal application of this method has been to social choice theory. We show that the use of this family of metrics generalizes the Kemeny method for aggregating the preferences in a population. We give examples illustrating the difference between the Kemeny method, our methods and other methods such as scoring rules. We then define two measures of conflict in a population. One is the measure of conflict with the social preference our method determines. The other is a measure of internal conflict in the population.

To illustrate the use of these conflict measures we then examine two further applications. One is to the selection of one explanation for an irrational choice function from among the many identified by Green and Hojman. The other is to modify and extend the measure of polarization defined by Esteban and Ray to the case of differences in preference rather than differences in income.

There are many open questions and further topics to be explored as a result of this paper. Here are a few of them.

1) Our paper has dealt primarily with the case of exchangeable distributions ν. When the decision problem is not permutation-invariant, new methods are required. They can be defined just as in this paper, but the computational techniques and most of our illustrative examples would need to be modified. Non-exchangeable problems are very realistic and constitute an important family of choice domains. They arise, for example, when the space of alternatives has a given algebraic structure, such as the product of different "issues" on which preferences can display cross-issue complementarity or substitutability, instead of being a set abstract alternatives that are treated symmetrically. In such a structure some feasible sets would arise only when a particular set of issues is "on the table". Random issues will not generate equal probabilities for all sets of the same cardinality.

Another example leading to non-exchangeability would arise when some alternatives are known in advance to be infeasible, or differentially less likely to arise. Disagreement about such alternatives should not influence the metric between preferences.

2) Our paper illustrates the importance of the frequency distribution of decision problems for the measurement of difference in preference. This suggests a new source of "data" for choice theory. Instead of focusing exclusively on the choice function c which describes the choice made at each feasible set, a theory of decision-making that takes into account the frequency of different opportunities could be constructed. In our terminology, the data of choice theory could be (c, ν) instead of only c. This would give us an empirical basis for selecting a particular metric from the family we have defined.

3) Our theory defines a mapping from Λ to Π corresponding to each distribution μ. Thus, given μ, the simplex Λ is divided into equivalence classes of populations that have the same social ordering. The algebraic structure of these partitions and their dependence on μ needs to be explored further. It is different from, but clearly related to, the structure of the equivalance classes defined by other rules such as the scoring rules studies by Saari[18].

APPENDIX

Axiomatic foundation for metrics on preference orderings

In the paper we define the metrics on Π as probabities of disagreement at a randomly chosen $A \in \mathcal{X}$. In this appendix we give a set of axioms that give rise to these metrics, without having to mention random feasible sets or to be specific about their distribution. Our axioms are very much like those of Kemeny and Snell (1962) and thus we will be brief. There is one key modification of the Kemeny-Snell axioms. It is this modification that allows for all the metrics in our family other than Kendall's tau.

We are concerned with a function

$$f : \Pi \times \Pi \to \mathbb{R}$$

Axiom 1 *f is a semi-metric*

That is, f is a non-negative, symmetric function satisfying the triangle inequality $f(\pi, \pi') + f(\pi', \pi'') \geqq f(\pi'', \pi)$

Axiom 2 *f is order-preserving under permutations*

If $f(\pi, \pi') \geqq f(\pi, \pi'')$ then for all permutations ρ, $f(\rho\circ\pi, \rho\circ\pi') \geqq f(\rho\circ\pi, \rho\circ\pi'')$

Axiom 3 *If π^i and π^j are obtained from e by a single adjacent tranposition at positions i and j respectively, then $i < j$ implies $f(\pi^i, e) \geqq f(\pi^j, e)$*

Axiom 4 *If $(\rho_0, \rho_1, ...\rho_M)$ is a minimal path from π to π' then $\sum_{k=1}^{M} f(\rho_{k-1}\rho_k) = f(\ \pi, \pi')$.*

Theorem 6 *Let $x(\pi)$ be the RBS signature of π and assume that f satisfies Axioms 1-4. There exists a non-increasing set of non-negative numbers d_k for $k = 2, ..., n$ such that*

$$f(\pi, e) = \sum_{k=2}^{n} x_k(\pi) d_k$$

The proof of this Theorem is exactly like the proof in Kemeny and Snell except that, because of the weakening of Axiom 4 to minimal paths instead of all paths (which Kemeny and Snell called "lines"), we can construct the distance from π to e only along minimal paths. By Theorem 6 above, all minimal paths have the same RBS signature. Thus, we need only specify the distances between two permutations that differ in a single transposition at i for all values of i. These can be any non-increasing numbers, according to Axiom 3. Then our Axiom 4 gives us the values of f as linear combinations of these distances. Because all minimal paths generate the same distance, f is well-defined.

Obviously Kendall's tau corresponds to $d_k = 1$ for all k. If $d_m = 0$ for any m then f will be a semi-metric but not a metric, as $f(\pi, \pi')$ can be zero for $\pi \neq \pi'$. When all $d_m > 0$ f will be a metric.

If $(\rho_0, \rho_1, ...\rho_M)$ is a path from π to π' but not a minimal path, as required by Axiom 4, then $\sum_{k=1}^{M} f(\rho_{k-1}\rho_k) \geqq f(\ \pi, \pi')$. To see this in an example, let us reconsider Example 1. There are two paths from $(a_3a_2a_1)$ to e given in this example: One is $(a_3a_2a_1)$, $(a_3a_1a_2)$, $(a_1a_3a_2)$, $(a_1a_2a_3)$. The other is $(a_3a_2a_1)$, $(a_2a_3a_1)$, $(a_2a_1a_3)$, $(a_1a_2a_3)$. The former is minimal while the latter is not. Calculating the distances along these paths, the former sums to $d_2 + 2d_3$ but the latter sums to $2d_2 + d_3$. Clearly these are equal in the case of the Kemeny metric and the triangle inequality holds strictly in all other cases. The minimal path defines the distance $f((a_3a_2a_1), e)$.

To find the exchangeable distribution on $\mathcal{X}$ corresponding to a given non-increasing vector $(d_2, ..., d_n)$ we can solve the system of equations derived from the formula in Theorem 4 for μ.

$$\begin{aligned} \text{For each } i &= 2, ..., n: \\ d_i &= \sum_{k=2}^{n} \frac{\mu_k}{\binom{n}{k}} \binom{n-i}{k-2} \end{aligned}$$

Proofs of Theorems

Proof of Theorem 1

Let $\pi \in \Pi$ and let ρ be the path generated by RBS as it sorts π into e. Assume $\pi \neq e$ and that $c_\pi(A) \neq c_e(A)$. Then $c_\pi(A)$ precedes $c_e(A)$ in the ordering π.

Each step in each pass of RBS involves a transposition of alternatives. We will say that y is promoted beyond x if the alternatives x and y are transposed at a step and x initially precedes y. As a pass of RBS proceeds, the alternatives that are promoted are continually better according to e. Thus in each pass of RBS there are three possibilities:

(i) There is some step in the pass where $c_e(A)$ is promoted and it continues to be promoted until it is promoted beyond $c_\pi(A)$.

(ii) There is some step where $c_e(A)$ is promoted but some $x \notin A$ becomes its immediate predecessor, $x < c_e(A)$, and $c_\pi(A)$ is a predecessor of x at this step. In this case $c_e(A)$ is no longer promoted during this pass. The next alternative to be promoted is then x or something better than x according to e. The pass will not end when $c_e(A)$ reaches x because something will be promoted beyond $c_\pi(A)$ on this pass. No member of A can be promoted beyond $c_\pi(A)$ in this pass.

(iii) There is no step where $c_e(A)$ is promoted. In this case there is some x that is promoted beyond $c_e(A)$. This x or something better than x according to e will be promoted beyond $c_\pi(A)$. No member of A can be promoted beyond $c_\pi(A)$ in this pass.

In case (i), at the step where $c_e(A)$ and $c_\pi(A)$ are transposed, the choice from A will change. By the definition of $c_e(A)$ it can never change again at any later step of the algorithm.

In cases (ii) and (iii) $c_\pi(A)$ will remain the highest-ranked member of A at all steps of this pass.

Thus on the first pass such that case (i) obtains there will be a change in the choice from A. At all future passes $c_e(A)$ is the highest ranking alternative among those in A and these passes will be in either case (ii) or case (iii). Thus the path generated by RBS is minimal. ■

Proof of Theorem 2

Let ρ and ρ' be two paths from π to e with RBS signatures x and x', respectively. Let k be the lowest index such that $x_k \neq x'_k$. The number of $A \in \mathcal{X}$ with cardinality $n-k+2$ at which the choice from A changes along these paths is determined by $(x_2, ..., x_k)$ and $(x'_2, ..., x'_k)$ because all transpositions at position $k+1$ and higher affect only sets of cardinality $n-k+1$ and smaller. Since $x_i = x'_i$ for all $i < k$, by the definition of k, these two paths cannot make the same number of changes at sets with cardinality $n-k+2$. Since both of these paths transform π into e, one of the paths must change one set of cardinality $n-k+2$ more than once. ■

Proof of Theorem 3

We need only verify the triangle inequality. This follows from the fact that the concatenation of minimal paths, one from π to π' and the other from π' to π'' may or may not be a minimal path from π to π''. If it is not minimal, strict inequality in the triangle inequality will hold.

As to the assertion that F is a metric whenever $\mu_2 \neq 0$, note that in this case there is a positive probability that A will be any particular pair. Thus if π and π' rank any pair differently, that pair could be the available set and the distance between these two orderings will be non-zero. (When $\mu_2 = 0$ then two orderings that differ only in their two lowest ranked alternatives will never disagree because some better alternative will be available and will be the choice of both.)■

Proof of Theorem 5

Define $\alpha(\pi, \lambda, A)$ to be the fraction of the population λ such that their most preferred element in A coincides with $c_\pi(A)$, the choice that would be made by the preference π. Consider the vector of $2^n - 1$ numbers $\alpha(\pi, \lambda, \cdot)$. If $\alpha(\pi, \lambda, \cdot)$ dominates each of the vectors $\alpha(\pi', \lambda, \cdot)$ for every $\pi' \in \Pi$, $\pi' = \pi$, then π will clearly be the assent-maximizing ordering. There is no tension in this case between assent-maximization and the requirment that social choice be generated by an ordering. On the other hand, if $\alpha(\pi, \lambda, A) < \alpha(\pi', \lambda, A)$ at any $A \in \mathcal{X}$, then if $\nu(A)$ is sufficiently large, π' will have a larger average assent than π.■

References

[1] Ambrus, Attila and Kareen Rozen. 2008. Revealed Conflicting Preferences.

[2] Apesteguia, Jose and Miguel A. Ballester. 2009. Choice by Sequential Procedures

[3] Arrow, Kenneth J. 1951. *Social Choice and Individual Values.* Wiley, New York

[4] Barthelemy, J.P. and B. Monjardet. 1981. The Median Procedure in Cluster Analysis and Social Choice Theory. *Mathematical Social Sciences, 1,* 235-268.

[5] Craven, John. 1996. Best Fit Social Choices: An Alternative to Arrow. *The Economic Journal 106(438), 1161-1174*

[6] deClippel, Geoffroy and Kfir Eliaz. 2009. Reason-Based Choice:A Bargaining Rationale for the Attraction and Compromise Effects.

[7] de Finetti, Bruno. 1931. Funzione caratteristica di un fenomeno aleatorio. Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale, 4, 251-299

[8] Diaconis, Persi. 1988. *Group Representations in Probability and Statistics.* Lecture Notes Monograph Series, volume 11. Hayward:Institute of Mathematical Statistics.

[9] Diaconis, Persi and David Freedman. 1980. Finite Exchangeable Sequences. Annals of Probability 8(4), 745-764

[10] Esteban, Joan-Maria and Debraj Ray. 1994. *Econometrica* 62(4), 819-851

[11] Green, Jerry R. and Daniel A. Hojman. 2008. Choice, Rationality and Welfare.

[12] Kalai, Gil, Ariel Rubinstein and Ran Spiegler. 2002. Rationalizing Choice Functions by Multiple Rationales, *Econometrica*, 70(6), 2481-2488

[13] Kemeny, John G. 1959. Mathematics Without Numbers. *Deadalus 88.* 577-591.

[14] Kemeny, John and J. Laurie Snell. 1962. Mathematical Models in the Social Sciences, chapter 2, Cambridge: MIT Press

[15] Klamler, Christian. 2008. A Distance Measure for Choice Functions. *Social Choice and Welfare 30, 419-425.*

[16] Moulin, Herve. 1988. Axioms of Cooperative Decision Making. Econometric Society Monograph number 15, Cambridge: Cambridge University Press

[17] Saari, Donald. 2001. Chaotic Elections. American Mathematical Society.

[18] Saari, Donald. 1994. The Geometry of Voting. Berlin, New York: Springer-Verlag

[19] Young, H. Peyton. 1988. Condorcet's Theory of Voting. *American Political Science Review. 82*(2). 1231-1244.

[20] Young, H. Peyton and Arthur Levenglick. 1978. A Consistent Extension of Condorcet's Election Principle, *SIAM Journal on Applied Mathematics, 35*(2), 285-300.