Is Sortition Both Representative and Fair?

Anonymous Author(s)
Affiliation
Address
email

Abstract

Sortition is a form of democracy built on random selection of representatives. Two of the key arguments in favor of sortition are *representation* (a random panel reflects the composition of the population) and *fairness* (everyone has a chance to participate). Uniformly random selection is perfectly fair, but is it representative? To answer this question, we introduce the notion of a representation metric on the space of individuals, and assume that the cost of an individual for a panel is determined by the q-th closest representative; the representation of a (random) panel is measured by the ratio between the (expected) sum of costs of the optimal panel for the individuals and that of the given panel. For $k/2 < q \leq k - \Omega(k)$, where k is the panel size, we show that uniform random selection is indeed representative by establishing a constant lower bound on this ratio. By contrast, for $q \leq k/2$, no random selection algorithm that is almost fair can give such a guarantee. We therefore consider relaxed fairness guarantees and develop a new random selection algorithm that sheds light on the tradeoff between representation and fairness.

1 Introduction

Most people think of democracy as synonymous with elections. But that has not always been the case: From the inception of democracy in ancient Athens until the American and French revolutions, democracy had typically been associated with random selection of representatives [1], a paradigm known as *sortition*.

Nowadays, sortition is mainly seen in the form of *citizens' assemblies* — randomly selected groups of people who deliberate on central questions, with the goal of informing policy. The impact and prevalence of citizens' assemblies around the world have motivated computational work on how to fairly and transparently select assembly members [2–4]. But there are signs that sortition is becoming even more widely accepted, including its recent institutionalization in Belgium, where permanent sortition-based bodies are now working alongside the parliaments of the German-speaking region and the Brussels region. In light of this progress, it may only be a matter of time until one of the many blueprints for sortition-based democracy [5] is implemented at the level of an entire country.

The excitement about sortition is driven by several appealing qualities, which are seen as providing solutions to some of the problems plaguing electoral democracy. We briefly present two of them in the context of *uniform selection*, which selects a uniformly random panel and is considered to be the ideal sortition method [6].

- *Descriptive representation*: A panel selected uniformly at random is likely to reflect the composition of the population from which it was drawn. Representation lends *legitimacy* to the process [7, 8], as citizens are able to identify some panelists who are similar to themselves.[1]

[1] Another argument for representation is epistemic: a diversity of opinions leads to better decisions [9].

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

- *Fairness*: Under uniform selection, each citizen has an equal chance to participate. Political theorists have argued that this quality realizes philosophical ideals like equality of opportunity and allocative justice [10].

By any reasonable measure of the fairness of selection probabilities — e.g., the minimum selection probability of any individual [2] — uniform selection achieves perfect fairness, as selection probabilities are equalized. We ask: *Is uniform selection also representative in a rigorous sense?* If we had a similar measure of representation, we would be able to evaluate whether this is the case. But quantifying representation poses a conceptual challenge.

Our approach. We address this challenge by assuming that there exists a *representation metric* on individuals, which measures to what degree one individual represents another (smaller distance means better representation). Readers familiar with the algorithmic fairness literature will no doubt make the connection to the *similarity metric* of Dwork et al. [11], which has been criticized on the grounds that it is difficult to explicitly construct [12]; a major obstacle is that the question of whether certain features should be used to determine similarity is domain-specific and tied to legal interpretation. By contrast, a representation metric is a more viable object, as it can be defined as a function of a common set of features that are routinely used by practitioners for this purpose, such as gender, age, ethnicity and education. Moreover, some of our main results, which pertain to uniform selection, are fully independent of the metric — for these it sufficies that such a metric *exists*.

Note, however, that a distance metric (on individuals) does not directly tell us to what degree an individual is represented by a panel. Following very recent work by Caragiannis et al. [13], we assume that the *cost* of a panel for an individual is determined by the q-th closest member of the panel, and our results are parameterized by q.

We can now define representation by taking a page from the literature on *distortion* in social choice [14]. Specifically, for a given selection algorithm, we measure its representation via the ratio between the social cost (sum of costs) of the optimal panel and that of the panel chosen by the algorithm, *in the worst case over underlying representation metrics*.

Our results. Returning to the question of whether uniform selection is also representative, and, more generally, the eponymous question of whether sortition is both representative and fair, our answer is that "it depends" — on the value of q.

When $k/2 < q \leq k - \Omega(k)$, where k is the size of the panel, we show that uniform selection (which is perfectly fair) achieves constant representation. Qualitatively, we view this as providing positive answers to our questions in the regime of $q > k/2$. Note that this regime has a natural interpretation: each individual wants a majority of the panel to be representative of themselves. This is especially justifiable when the panel makes decisions or recommendations through voting, which is often the case in citizens' assemblies.

By contrast, for the regime of $q \leq k/2$, we prove that any selection algorithm that chooses each individual with probability somewhat higher than q/n, where n is the number of individuals, must have representation of precisely 0. This result clearly applies to uniform selection, where the minimum selection probability is k/n, and it motivates us to consider weaker fairness guarantees. We design an algorithm, RANDOMREPLACE, which selects each individual with probability at least q/n and has a nontrivial representation guarantee of $1/(q+1)$ for any value of q.

Lastly, we run experiments to compare the different selection algorithms using two real datasets. The experiments show that worst-case representation guarantees predict representation in practice, and give us a more nuanced understanding of the performance of different selection algorithms in terms of representation.

Related work. The design of practical, fair and transparent algorithms for selecting citizens' assemblies was explored in several previous papers [2–4] — two of which appeared in previous NeurIPS conferences. Assemblies are required to be representative of the population with respect to features like gender, age, ethnicity, education and geography. This is done by setting quotas on individual features; for example, a panel of 100 people might be required to include at least 48 men, at least 48 women, and at least 2 people who identify as non-binary. The challenge is that an assembly is selected from a pool of volunteers, which is typically unrepresentative of the population due to self-selection bias. Uniformly random selection, therefore, would likely itself result in an

unrepresentative panel. Instead, the primary selection algorithm advocated by Flanigan et al. [2] computes a distribution over quota-compliant panels that (roughly speaking) maximizes the minimum selection probability of any volunteer, thereby maximizing fairness subject to the demographic constraints. By contrast, like Benadè et al. [15] and the political theory literature, we take a longer-term view: We are interested in random selection directly from the population, which is a hallmark of some plans for sortition-based democracy [5]. In addition, we take a fundamentally different, and arguably more nuanced, view of representation. Indeed, our framework can accommodate questions of intersectionality (to what degree is a rural, college-educated man represented by a rural, college-educated woman?) and also provides a more holistic analysis of the composition of the panel (to what degree is a rural, college-educated man represented by a panel that includes 10 rural, college-educated men and 99 urban women with no college education?).

Our approach to evaluating representation through a metric is rooted in spatial theories of voting from the political theory literature [16, 17]. The idea of measuring how poorly a panel represents an individual by the distance of the q-th closest panel member to the individual was introduced by Caragiannis et al. [13] in the context of committee elections. Finally, aiming to minimize the total misrepresentation to all people and comparing that to the optimal panel makes our representation measure the inverse of *distortion* in voting theory [18, 19], where distortion is the inverse of our representation measure. In voting, it is assumed that we only have partial access to the metric in the form of voters' ranked preferences over the candidates induced by distance comparisons. In contrast, our results for uniform selection use no knowledge of the underlying metric, while our other algorithms assume complete access. When selecting a single candidate as the winner, it is known that the best distortion achievable by deterministic selection is 3 (which maps to $1/3$ representation in our formulation) [20]. Our setting is closer to committee selection, where a committee of k candidates is selected. Here, Caragiannis et al. [13] show that when each voter measures her distance to the q-th closest committee member, there is a trichotomy: the best possible distortion is infinite when $q \leq k/3$, linear in the number of voters when $q \in (k/3, k/2]$, and 3 for deterministic selection when $q > k/2$. Sortition is a special case of committee elections in which the set of candidates is the same as the set of voters. Hence, all the positive results from Caragiannis et al. [13] carry over in the absence of any fairness constraints. However, our results show that when (perfect) fairness in selection probabilities is sought in conjunction with representation, the distortion becomes infinite (zero representation) for all $q \leq k/2$ but constant distortion can still be achieved for $q > k/2$. The idea of the set of voters acting as the set of candidates was explored by Cheng et al. [21, 22]. However, they model infinitely many voters using a continuous distribution over the metric space.

2 Preliminaries

For all $t \in \mathbb{N}$, define $[t] = \{1, \ldots, t\}$. Let $N = [n]$ be the set that constitutes the underlying population. A *panel* P is a subset of the population. Let $\mathcal{S}_k(N)$ denote the set of all subsets of N of size k. We omit N when it is clear form the context. The population lies in an underlying metric space endowed with distance d, which we think of as the *representation metric* discussed earlier. For each $i, j \in N$, $d(i, j)$ denotes the distance between i and the following properties are satisfied: (a) $d(i, j) \geq 0$, and $d(i, j) = 0$ if and only if $i = j$, (b) $d(i, j) = d(j, i)$, and (c) for each $i, j, \ell \in N$, $d(i, \ell) + d(\ell, j) \geq d(i, j)$. The last property is known as the triangle inequality. An instance of our problem is given by the underlying population along with distances as defined by d; hereinafter, we simply denote such an instance by d.

Given a panel P of size k and a positive integer $q \in [k]$, the *q-cost* of individual i for P, denoted by $c_q(i, P; d)$, is equal to the distance of i from her q-th closest representative in P. Note that for $q = 1$, we have $c_1(i, P; d) = \min_{j \in P} d(i, j)$ and for $q = k$ we get $c_k(i, P; d) = \max_{j \in P} d(i, j)$. Let $\mathsf{top}_q(i, P; d)$ be the set of q closest members of P to i (ties broken arbitrarily). The *q-social cost* of panel P is given by $\mathrm{SC}_q(P; d) = \sum_{i \in N} c_q(i, P; d)$, i.e., the sum of the q-costs over all individuals. We omit d from the notation when it is clear from the context.

In this setting, a *selection algorithm* $\mathcal{A}_{k,q}$ that is defined over k and q takes as input d and outputs a distribution over all panels of size k. We are especially interested in the *uniform selection* algorithm, denoted by $\mathcal{U}_k$, that always outputs a uniform distribution over $\mathcal{S}_k$, independently of the value of q. In other words, it does not take into account the underlying metric space or q, but instead outputs a committee of size k chosen uniformly at random.

Fairness: As discussed in the introduction, the main property of uniform selection is that each individual is selected to be part of the panel with probability exactly equal to k/n, i.e., $\Pr[i \in \mathcal{U}_k] = \frac{k}{n}$. In other words, all the individuals are treated equally since they have the same chances to be chosen. We call this property *perfect fairness*, and in general an algorithm $\mathcal{A}_{k,q}$ provides perfect fairness when for each instance d, it ensures that $\min_{i \in N} \Pr[i \in \mathcal{A}_{k,q}] = k/n$.

When perfect presentation is too restrictive, we relax this constraint by allowing some individuals to be selected with probability less than k/n. Then, we measure the fairness of an algorithm $A_{k,q}$ as the worst-case ratio of the minimum probability of an individual to be selected by the algorithm and the ideal selection probability k/n. More formally,

$$\mathsf{fairness}_q(\mathcal{A}_{k,q}) = \inf_d \frac{\min_{i \in N} \Pr[i \in \mathcal{A}_{k,q}(d)]}{k/n}.$$

Representation: As we have discussed above, another key property we are interested in measuring is representation. To do so, we consider a panel to be a good representative of the whole population when the q-social cost is not much larger than the optimum. In other words, a selection algorithm $\mathcal{A}_{k,q}$ that outputs a distribution over the different committees of size k provides good representation when the expected q-social cost of the panel is similarly small. More formally, we define the *representation* of a selection algorithm $\mathcal{A}_{k,q}$ as the worst-case ratio of the minimum possible social q-cost of any panel and the expected q-social cost of the panel chosen by $\mathcal{A}_{k,q}$ over all possible instances, i.e.,

$$\mathsf{repr}_q(\mathcal{A}_{k,q}) = \inf_d \frac{\min_{P' \in \mathcal{S}_k(N)} \mathrm{SC}_q(P'; d)}{\mathbb{E}[\mathrm{SC}_q(\mathcal{A}_{k,q}(d)]}.$$

3 Representation with Perfect Fairness for $q > k/2$

We start with the case that $q > k/2$. We show that in this case uniform selection is asymptotically optimal with respect to representation among all selection algorithms that are perfectly fair. Moreover, the representation of uniform selection is constant for any $q = c \cdot k$ for $1/2 < c < 1$.

Theorem 1. *For $q > k/2$, uniform selection satisfies $\mathsf{repr}_q(\mathcal{U}_k) \geq \frac{1}{2} \cdot \frac{k-q+1}{k}$.*

A crucial property that we exploit in this section is that the q-costs of the individuals satisfy the triangle inequality when $q > k/2$. This observation was first made by Caragiannis et al. [13]; we present the lemma below for completeness.

Lemma 1. *For $q > k/2$, individuals $i, j \in N$, and a panel P, $c_q(i, P; d) + c_q(j, P; d) \geq d(i, j)$.*

Proof. Let $T_i = \mathsf{top}_q(i, P)$ and $T_j = \mathsf{top}_q(i, P)$ be the q closest neighbors of i and j, respectively, in the panel P. As $|T_i| = |T_j| > k/2$, there exists an individual $k \in T_i \cap T_j$. Therefore,

$$d(i, j) \leq d(i, k) + d(k, j) \leq c_q(i, P) + c_q(j, P). \qquad \square$$

This observation enables us to show the following lower bound on the social cost of the optimal committee.

Lemma 2. *For $q > k/2$, the q-social cost of the optimal panel P^* is at least*

$$\mathrm{SC}_q(P^*; d) \geq \frac{1}{2(n-1)} \cdot \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i, j).$$

Proof. By applying Lemma 1 for all pairs of individuals (i, j), we get

$$\sum_{i \in N} \sum_{j \in N \setminus \{i\}} \Big(c_q(i, P^*) + c_q(j, P^*) \Big) \geq \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i, j).$$

The q-cost of each person appears exactly $2(n-1)$ times on the left hand side. Thus,

$$\mathrm{SC}_q(P^*; d) = \sum_{i \in N} c_q(i, P^*) \geq \frac{1}{2(n-1)} \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i, j). \qquad \square$$

We are now ready to prove the theorem.

Proof of Theorem 1. For any committee P of size k,

$$c_q(i,P) = \min_{q' \in \{q,\dots,k\}} c_{q'}(i,P) \le \frac{1}{k-q+1} \sum_{q' \in [q,k]} c_{q'}(i,P)$$
$$\le \frac{1}{k-q+1} \sum_{q' \in [1,k]} c_{q'}(i,P) = \frac{1}{k-q+1} \sum_{j \in P} d(i,j),$$

where in the first inequality the minimum is upper bounded by the average. Therefore, the expected social cost of uniform selection is at most

$$\begin{aligned}
\mathbb{E}[\mathrm{SC}_q(\mathcal{U}_k(N))] &= \sum_{i \in N} \mathbb{E}_{P \sim \mathcal{U}_k}\left[c_q(i,P)\right] \\
&\le \frac{1}{k-q+1} \sum_{i \in N} \mathbb{E}_{P \sim \mathcal{U}_k}\Big[\sum_{j \in P} d(i,j)\Big] \\
&= \frac{1}{k-q+1} \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i,j) \cdot \Pr_{P \sim \mathcal{U}_k}[j \in P] \\
&= \frac{1}{k-q+1} \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i,j) \cdot \frac{k}{n}.
\end{aligned}$$

By Lemma 2 and the upper bound shown above, we have

$$\mathsf{repr}_q(\mathcal{U}_k) \ge \frac{\frac{1}{2(n-1)} \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i,j)}{\frac{1}{k-q+1} \cdot \frac{k}{n} \cdot \sum_{i \in N} \sum_{j \in N \setminus \{i\}} d(i,j)} = \frac{1}{2} \cdot \frac{n}{n-1} \cdot \frac{k-q+1}{k} \ge \frac{1}{2} \cdot \frac{k-q+1}{k}. \quad \square$$

In the proof above, the only property of uniform selection we use is that the marginal probabilities are equal to k/n. Hence, this lower bound also holds for any perfectly fair selection algorithm.

We next establish an upper bound on the representation of any perfectly fair selection algorithm. It shows that the lower bound of Theorem 1 is tight up to a factor of 4.

Theorem 2. *For any $q > k/2$, every selection algorithm $\mathcal{A}_{k,q}$ with $\mathsf{fairness}(\mathcal{A}_{k,q}) = 1$ satisfies $\mathsf{repr}_q(\mathcal{A}_{k,q}) \le 2 \cdot \frac{k-q+1}{k+1}$.*

Proof. First, note that if $q = \frac{k+1}{2}$, the statement trivially holds, since $\mathsf{repr}_q(\mathcal{A}_{k,q}) \le 2 \cdot \frac{k-q+1}{k+1} = 1$ which is true for any algorithm. Thus, we assume $q > \frac{k+1}{2}$. Consider an instance with $n = k+1$ individuals where $k-q+1$ individuals are located at 0 and q individuals are at 1, denoted by N_0 and N_1, respectively. Any committee of size k leaves one person out of the committee, and for each individual $i \in N$, this happens with probability of

$$\Pr_{P \sim \mathcal{A}_{k,q}}[i \notin P] = 1 - \Pr_{P \sim \mathcal{A}_{k,q}}[i \in P] = (1 - \tfrac{k}{k+1}) = \tfrac{1}{k+1}.$$

Individuals in N_0 will always have a q-cost of 1, because $k+1-q < \frac{k+1}{2} < q$ individuals are located there. Therefore, $\mathbb{E}_{P \sim \mathcal{A}_{k,q}}[\sum_{i \in N_0} c_q(i,P)] = |N_0|$. For individuals in N_1, their q-cost is 1 if and only if less than q individuals are selected from N_1, i.e., the single person left out of the committee is located at 1. This event happens with probability

$$\Pr_{P \sim \mathcal{A}_{k,q}}[\textstyle\bigcup_{i \in N_1} (i \notin P)] = \sum_{i \in N_1} \Pr_{P \sim \mathcal{A}_{k,q}}[i \notin P] = \frac{|N_1|}{k+1},$$

where the first equality comes from the fact that the events are disjoint (which holds because any committee leaves out exactly one individual). Therefore, $\mathbb{E}_{P \sim \mathcal{A}_{k,q}}[\sum_{i \in N_1} c_q(i,P)] = |N_1| \cdot \frac{|N_1|}{k+1}$, and the expected social cost of any perfectly fair algorithm is

$$\mathbb{E}_{P \sim \mathcal{A}_{k,q}}[\mathrm{SC}_q(P;d)] = |N_0| + |N_1| \cdot \frac{q}{k+1} \ge |N_0| + |N_1| \cdot \frac{1}{2}.$$

The optimal committee would leave out a person from N_0 and achieve a social cost of $|N_0|$. Therefore, the representation of any algorithm with perfect fairness is at most

$$\frac{|N_0|}{|N_0| + \frac{1}{2} \cdot |N_1|} \le \frac{|N_0|}{\frac{1}{2} \cdot |N_0| + \frac{1}{2} \cdot |N_1|} = 2 \cdot \frac{k-q+1}{k+1}. \quad \square$$

ALGORITHM 1: RANDOMREPLACE$_{k,q}$

1: Compute an optimal panel $P^* \in \arg\min_{P \in \mathcal{S}_k} \mathrm{SC}_q(P)$
2: Pick $S \in \mathcal{S}_q$ uniformly at random
3: Set $P_S \leftarrow P^*$
4: **for** $i \in S \setminus P_S$ **do**
5: Pick an arbitrary $j_i \in \mathsf{top}_q(i, P^*) \setminus S$
6: $P_S \leftarrow P_S \cup \{i\} \setminus \{j_i\}$
7: **end for**
8: **return** P_S

4 Representation with Relaxed Fairness for $q \leq k/2$

In stark contrast to the case of $q > k/2$, uniform selection and, more generally, any perfectly fair selection algorithm, cannot obtain bounded representation when $q \leq k/2$. In fact, the following theorem shows that selection algorithms with fairness strictly more than $\frac{q+(k \bmod q)}{k}$ (which itself is upper bounded by $(2q-1)/k$) suffer from unbounded representation. The proof is in Appendix A.

Theorem 3. *For $q \leq k/2$, $\epsilon > 0$, and any selection algorithm $\mathcal{A}_{k,q}$ with $\mathsf{fairness}(\mathcal{A}_{k,q}) \geq \frac{q+(k \bmod q)}{k} + \epsilon$, $\mathsf{repr}_q(\mathcal{A}_{k,q})$ is 0.*

Although bounded representation is not feasible with fairness slightly larger than $\frac{q}{k}$, we design a selection algorithm that can achieve representation of $q+1$ with fairness of $\frac{q}{k}$. RANDOMREPLACE$_{k,q}$, in Algorithm 1, starts with an optimal panel P^*, randomly selects a panel S of q individuals, and replaces individuals in P^* with individuals in S as follows. First, individuals in $S \cap P^*$ remain in the final panel. Then, for $i \in S \setminus P^*$, swap i with one of its q-closest neighbors in the optimal panel $\mathsf{top}_q(i, P^*)$ that has not been replaced by the algorithm yet. The next theorem establishes the fairness and representation guarantees of RANDOMREPLACE.

Theorem 4. *For any $q \in [k]$, we have that $\mathsf{repr}_q(\text{RANDOMREPLACE}_{k,q}) \geq \frac{1}{q+1}$ and $\mathsf{fairness}(\text{RANDOMREPLACE}_{k,q}) \geq \frac{q}{k}$.*

Proof. Let P^* be an optimal panel and $S \subseteq N$ be a set of size q chosen uniformly at random. We denote with P_S the panel that is returned from the algorithm. First, we show that Line 5 of the algorithm is valid. The algorithm reaches this line when considers $i \in S \setminus P_S$, meaning that i is not included in the panel P_S and therefore is not included in P^*. Hence, $\mathsf{top}_q(i, P^*) \setminus S$ cannot be empty since $|\mathsf{top}_q(i, P^*)| = q$, $|S| = q$ and there exists i in S but not in $\mathsf{top}_q(i, P^*)$.

We see that that every individual in S is included in P_S as in Line 6 the algorithms ensures that each such individuals is included and is never excluded afterwards. Hence, as each individual is chosen in S with probability at least q/n, we can see that $\mathsf{fairness}(\text{RANDOMREPLACE}_{k,q}) \geq q/k$.

Now, we prove that for any $S \in \mathcal{S}_q$ and any agent $i' \in N$,

$$c_q(i', P_S) \leq c_q(i', P^*) + \max_{i \in S} c_q(i, P^*) \leq c_q(i', P^*) + \sum_{i \in S} c_q(i, P^*). \tag{1}$$

The second inequality holds as the maximum is at most the sum. Therefore, we focus on the first. If $c_q(i', P_S) \leq c_q(i', P^*)$, then it trivially holds. Hence, assume that $c_q(i', P_S) > c_q(i', P^*)$. In this case, we can show that there exists $i \in S \setminus P^*$ such that $d(i', i) \geq c_q(i', P_S)$ and $j_i \in \mathsf{top}_q(i', P^*)$. First, note if for each $i \in S \setminus P^*$, j_i does not belong in $\mathsf{top}_q(i', P^*)$, then it is not possible that $c_q(i', P_S) > c_q(i', P^*)$, since $\mathsf{top}_q(i', P^*) \subseteq P_S$. Next, suppose for contradiction that for every i that was included in P_S when $j_i \in \mathsf{top}_q(i', P^*)$ was excluded from it, it holds that $d(i', i) < c_q(i', P^*)$. Then, in P_S there are $|\mathsf{top}_q(i', P^*) \setminus \cup_{i \in S \setminus P^*}(\{j_i\} \cap \mathsf{top}_q(i', P^*))|$ individuals that have distance at most $c_q(i', P^*) < c_q(i', P_S)$ from i' and $|\{i \in S \setminus P^* : j_i \in \mathsf{top}_q(i', P^*)\}|$ individuals that have distance less than $c_q(i', P_S)$ from i'. Note that

$$|\cup_{i \in S \setminus P^*} (\{j_i\} \cap \mathsf{top}_q(i', P^*))| = |\{i \in S \setminus P^* : j_i \in \mathsf{top}_q(i', P^*)\}|,$$

and hence we get that there are at least $|\mathsf{top}_q(i', P^*)| = q$ individuals in P_s with distance strictly less than $c_q(i', P_S)$ from i' and we reach a contradiction.

From the above observation , we have that

$$c_q(i', P_S) \leq d(i', i) \leq d(i', j_i) + d(j_i, i) \leq c_q(i', P^*) + c_q(i, P^*) \leq c_q(i', P^*) + \max_{i \in S} c_q(i, P^*)$$

where the penultimate inequality holds because $j_i \in \mathsf{top}_q(i', P^*)$ and $j_i \in \mathsf{top}_q(i, P^*)$ form the definition of j_i. This proves (1).

Summing (1) over all $i' \in N$, we have

$$\mathrm{SC}_q(P_S) \leq \mathrm{SC}_q(P^*) + n \cdot \sum_{i \in S} c_q(i, P^*).$$

Taking the expectation of the above equation with respect to S, and using the fact that $\Pr[i \in S] = q/n$, we have

$$\mathbb{E}[\mathrm{SC}_q(P_S)] \leq \mathrm{SC}_q(P^*) + n \cdot \sum_{i \in N} \frac{q}{n} \cdot c_q(i, P^*) = (q+1) \cdot \mathrm{SC}_q(P^*),$$

as needed. □

The next theorem shows that when $q = \Omega(k)$, RANDOMREPLACE$_{k,q}$ attains an asymptotically optimal fairness-representation tradeoff. The proof is in Appendix A.

Theorem 5. *For any $q \leq k/2$ such that $k \bmod q = 0$, every selection algorithm $\mathcal{A}_{k,q}$ with* $\mathsf{fairness}(\mathcal{A}_{k,q}) \geq q/k$ *satisfies* $\mathsf{repr}_q(\mathcal{A}_{k,q}) \leq k/q^2$.

From the above theorem, it follows that Algorithm 1 achieves the highest possible fairness of q/k subject to positive representation.

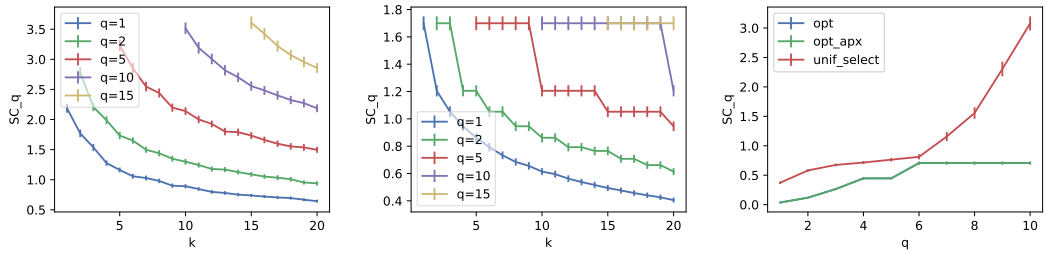
5 Experiments

Next, we conduct an empirical comparison between the selection algorithms that we have considered above. Data on the metric-structure preferences of groups in their full richness are difficult to come by, but it is reasonable to expect that the extent to which individuals feel well-represented by one another is at least partly a function of their relative characteristics along some observable features.

We therefore begin with two datasets that express the characteristics of populations along a range of features, and randomly construct synthetic metric preferences from these feature signatures. We think it reasonable to expect that the resulting metrics capture the high-dimensional correlations between salient features which may influence individuals' positions in the metric space and the variety of opinion within populations, even if they stop short of encoding the population's true preferences for any particular issue of interest.

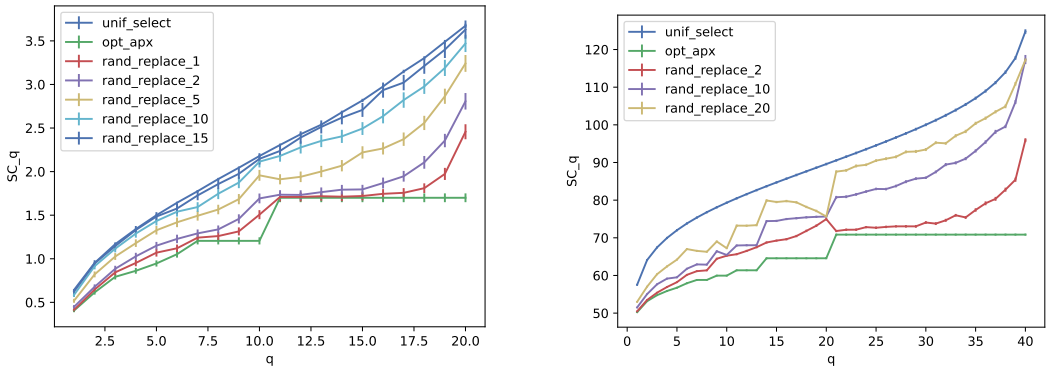
Metric construction. We choose some set of features along which to evaluate the individuals in the population. These are the features which will inform our metric. For each feature s, F^s is the set of possible values that this feature can take. Each individual i is then represented as a vector of feature values, where f_i^s is the value that i has for feature s. For each feature we sample a weight $w_t \sim U[0,1]$ uniformly at random from the interval $[0,1]$. Finally the distance between individuals i and j and our metric d is defined to be $d(i,j) := \sum_s w_s \cdot (1 - \mathbb{1}\{f_i^s = f_j^s\})$.

Data sets and experimental parameters. Our first source of demographic data is the UCI `Adult` dataset, which was derived from the 1994 Current Population Survey of the US Census Bureau, and is made available by the UCI Machine Learning Repository under a CC BY 4.0 license [23]. It contains a range of demographic variables principally related to employment. Our experiments do not require `Adult` to be representative of any actual population, nor should this an assumption be made lightly [24]. For `Adult` we choose the features `workclass`, `education`, `marital status` and `sex`. `Adult` contains $n = 30162$ individuals with values for each feature, who may be viewed as a distribution over the 721 unique feature vectors which they collectively hold. For the empirical evaluation of our selection algorithms on these randomly generated metrics, we suppose that this is in fact the distribution for a population large enough that there are at least k individuals with any given feature vector. As these metrics represent populations, the q-social costs in Figure 1 and Figure 2 are



(a) UNIFORMSELECTION (b) OPTPROXY (c) Comparison to $SC_q(P^*, d)$ on small instances ($k = 10$).

Figure 1: The q-social cost of UNIFORMSELECTION and OPTPROXY for $k = 1$ to $k = 20$ and a selection of q, based on `Adult`.



(a) `Adult` with $k = 20$. (b) ESS UK with $k = 40$.

Figure 2: Comparison of different algorithm for fixed k, where RANDOMREPLACE$_{r,q}$ is applied to the panel selected by OPTPROXY. As r ranges from 0 to k, the q-social cost of RANDOMREPLACE$_{r,q}$ interpolates between that of OPTPROXY and UNIFORMSELECTION.

normalized by population size. Figure 1 and Figure 2a depict data averaged over 100 random metrics constructed in this manner.

Our second source of demographic data is the *European Social Survey* (ESS) [25], which is made available by the Norwegian Centre for Research Data under a CC BY 4.0 license. We use the ESS Round 9 (2018) data, which consists of 46,276 people in 27 countries, and contains ∼1450 features regarding socioeconomic demographic, political beliefs, geographical region, house-hold composition, personal values, media use and trust, etc. Most of the features are country-specific, which leaves roughly 250 features available per country, while each country has between 781 and 2745 entries (with a mean of 1713). Each entry is assigned an analysis weight which is aimed to correct the differential selection probabilities. In contrast to our experiments with `Adult`, we use all of the available features available in ESS. The metric construction is similar to `Adult`. Figure 2b is based on the data of the United Kingdom (2204 entries), averaged over 100 randomly constructed metrics. Similar to `Adult`, we assume there are at least k people associated with each feature vector.

In evaluating UNIFORMSELECTION and RANDOMREPLACE, we use a proxy for the optimal q-social cost, since n and k are too large to support finding $SC_q(P^*, d)$ exactly. This selection algorithm OPTPROXY is an implementation of the fault-tolerant metric k-medians algorithm of Kumar et al. [26], which guarantees a constant-factor approximation to $SC_q(P^*, d)$. This algorithm uses a constant-factor metric k-medians algorithm as a primitive; we implement the local search algorithm of Arya et al. [27] with single swaps. In order to evaluate the fidelity of OPTPROXY we compare it with $SC_q(P^*, d)$ for 100 metrics d constructed by drawing 30 randomly chosen feature vectors from the supports of `Adult` instances described above. For panels of size $k = 10$ and all values of q we find that OPTPROXY recovers $SC_q(P^*, d)$ exactly (Figure 1c).

Experimental Results. In Figure 1a, we see UNIFORMSELECTION behaving as expected. For a fixed k, the q-social cost of a uniformly random panel is reliably higher for larger q, and for a fixed q, it decays smoothly as we increase the panel size k. The behavior of OPTPROXY under the same conditions (fig. 1b) is rather choppier. Here, all of our approximately optimal panels P_q start at the same value of $SC_q(P_q, d)$ when $q = k$ and decay as q remains fixed and the size of the panel k increases. This first property is to be expected: when $q = k$ and points in the metric may appear on the panel multiple times (as is the case with these distributions), it is without loss of generality to take the optimal panel P_q^* to be k copies of a (carefully chosen) single point in the metric. For these panels P_q, OPTPROXY is simply finding that point.

The fact that the q-social cost attained by OPTPROXY for fixed q and $k/2 < q < k$ appears constant, on the other hand, is not universally true of $SC_q(P^*, d)$. These plateaus are due to the way OPTPROXY selects panels. For this range of q, OPTPROXY selects q individuals from the optimal 1-median location and chooses the remaining $k - q$ uniformly at random. Hence, when $q > k/2$, it is guaranteed that the qth closest member in the panel is at this optimal 1-median location. More generally, OPTPROXY selects q individuals from an (approximately) optimal $\lfloor k/q \rfloor$-median solution, which explains the other steps. Since P_q^* need not be of this form, it is interesting to note that it exhibits this same step-like behavior in Figure 1c.

Finally, we consider Figure 2a and Figure 2b, which illustrate the q-social costs of RANDOMREPLACE$_{r,q}$ for $k = 20$ and for $k = 40$, of the `Adult` and `ESS` data, respectively, and for a range of r and q. OPTPROXY again exhibits step-like behavior, consistent with Figure 1b. On the other hand, UNIFORMSELECTION increases very smoothly from $q = 1$ to $q = k$, which is suggestive of a metric that is not comprised of a few well-separated groups. Additionally, we see RANDOMREPLACE$_{r,q}$ interpolating nicely from OPTPROXY at $q = 0$ up to UNIFORMSELECTION up at $q = k$. The strangest portion of this plot is the kink that appears in RANDOMREPLACE$_{q/2,q}$ at $q = k/2$. This inversion is perhaps odd because we expect $SC_q(P, d)$ to be monotonically increasing in q for any fixed panel P. This may be due to the implementation of PROXYOPT: if $q = k/2$ then PROXYOPT chooses $k/2$ individuals from each of the optimal 2-medians centers. After replacing some of this panel, every individual is now likely closest to either some random other individual or the further center. In the likely event that these two locations are far apart, we should expect this replacement to dramatically increase $SC_{q/2}(P, d)$ for the average individual. We can perhaps view this as providing an illustration of the inapproximability of $\mathsf{repr}_q(P^*)$ by UNIFORMSELECTION, in miniature: if these two 2-medians centers were the entirety of their metric, then any swaps at all would send the expected $q = k/2$-social cost from zero to nonzero.

6 Discussion

Our results in Section 4 show that in some cases, relaxing fairness requirements allows improving representation dramatically. More generally, it is interesting to understand the tradeoff between representation and fairness, and to chart the Pareto frontier. In Appendix B, we take some first steps in this direction. One observation is that a generalization of RANDOMREPLACE that replaces r individuals instead of q gives representation at least $1/(r+1)$ and fairness at least r/k. We also show that a random-dictatorship-like algorithm gives nontrivial fairness and representation guarantees in the regime of $q > k/2$. However, there remain several open questions: for example, when $q \leq k/2$, what level of fairness can be achieved if we seek constant representation?

We focused our attention on the q-cost formulation of Caragiannis et al. [13], in which each individual measures their distance to the q-th closest panel member. One can analyze the representation-fairness tradeoff with other cost functions. For example, what if different individuals use different values of q? Another appealing choice is when each individual measures the *average* distance to all panel members; in Appendix C, we show that for this cost function, uniform selection achieves representation of at least $1/2$ for all $q > k/2$.

More broadly, one can use measures of representation other than the *total cost* to all individuals. For example, one may wish to use a selection algorithm that minimizes the *maximum cost* to any individual, or strikes a balance between maximum cost and total cost. We hope that answers to some of these questions will lead to a better understanding of the strengths of sortition, and to new ways of realizing this democratic paradigm.

References

[1] D. Van Reybrouck. *Against Elections: The Case for Democracy*. Random House, 2016.

[2] B. Flanigan, P. Gölz, A. Gupta, B. Hennig, and A. D. Procaccia. Fair algorithms for selecting citizens' assemblies. *Nature*, 596:548–552, 2021.

[3] B. Flanigan, P. Gölz, A. Gupta, and A. D. Procaccia. Neutralizing self-selection bias in sampling for sortition. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[4] B. Flanigan, G. Kehne, and A. D. Procaccia. Fair sortition made transparent. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[5] J. Gastil and E. O. Wright, editors. *Legislature by Lot: Transformative Designs for Deliberative Governance*. Verso, 2019.

[6] F. Engelstad. The assignment of political office by lot. *Social Science Information*, 28(1):23–50, 1989.

[7] J. M. Parker. *Randomness and Legitimacy in Selecting Democratic Representatives*. PhD thesis, University of Texas at Austin, 2011.

[8] J. Fishkin. *Democracy When the People Are Thinking: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press, 2018.

[9] H. Landemore. Deliberation, cognitive diversity, and democratic inclusiveness: An epistemic argument for the random selection of representatives. *Synthese*, 190:1209–1231, 2013.

[10] P. Stone. *The Luck of the Draw: The Role of Lotteries in Decision Making*. Oxford University Press, 2011.

[11] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. S. Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*, pages 214–226, 2012.

[12] M.-F. Balcan, T. Dick, R. Noothigattu, and A. D. Procaccia. Envy-free classification. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 1238–1248, 2019.

[13] I. Caragiannis, N. Shah, and A. A. Voudouris. The metric distortion of multiwinner voting. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 2022. Forthcoming.

[14] E. Anshelevich, A. Filos-Ratsikas, N. Shah, and A. A. Voudouris. Distortion in social choice problems: The first 15 years and beyond. arXiv:2103.00911, 2021.

[15] G. Benadè, P. Gölz, and A. D. Procaccia. No stratification without representation. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC)*, pages 281–314, 2019.

[16] James M Enelow and Melvin J Hinich. *The spatial theory of voting: An introduction*. CUP Archive, 1984.

[17] Kenneth Arrow. *Advances in the spatial theory of voting*. Cambridge University Press, 1990.

[18] A. D. Procaccia and J. S. Rosenschein. The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*, pages 317–331, 2006.

[19] E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27–51, 2018.

[20] V. Gkatzelis, D. Halpern, and N. Shah. Resolving the optimal metric distortion conjecture. In *Proceedings of the 61st Symposium on Foundations of Computer Science (FOCS)*, pages 1427–1438, 2020.

[21] Y. Cheng, S. Dughmi, and D. Kempe. Of the people: voting is more effective with representative candidates. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*, pages 305–322, 2017.

[22] Y. Cheng, S. Dughmi, and D. Kempe. On the distortion of voting with multiple representative candidates. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 973–980, 2018.

[23] D. Dua and C. Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

[24] F. Ding, M. Hardt, J. Miller, and L. Schmidt. Retiring Adult: New datasets for fair machine learning. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 6478–6490, 2021.

[25] ESS Round 9: European Social Survey (2021): ESS-9 2018 Documentation Report. Edition 3.1. bergen, european social survey data archive, nsd - norwegian centre for research data for ESS ERIC. 2021. doi: 10.21338/NSD-ESS9-2018. URL https://www.europeansocialsurvey.org/data/.

[26] N. Kumar and B. Raichel. Fault tolerant clustering revisited. In *Proceedings of the 25th Canadian Conference on Computational Geometry, CCCG 2013, Waterloo, Ontario, Canada, August 8-10, 2013*. Carleton University, Ottawa, Canada, 2013.

[27] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *SIAM J. Comput.*, 33(3):544–562, 2004.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]