DISTANCE RATIONALIZATION OF ANONYMOUS AND HOMOGENEOUS VOTING RULES

BENJAMIN HADJIBEYLI AND MARK C. WILSON

Abstract. The concept of distance rationalizability of voting rules has been explored in recent years by several authors. All previous work has dealt with a definition in terms of preference profiles. However most voting rules in common use are anonymous and most are also homogeneous. In this case there is a much more succinct representation (using the voting simplex) of the rule. This representation has been widely used in the voting literature, but not in the context of distance rationalizability.

We first define distance rationalizability in this new framework and explain in detail the connection to the original definition. In doing so we unify, correct, and extend previous work. The simplex interpretation yields a natural connection to areas of continuous mathematics not seen before in the voting literature, namely Wasserstein spaces and Minkowski spaces.

We prove some positive and some negative results about the decisiveness of distance rationalizable rules on the simplex. The positive results connect with the recently developed theory of hyperplane rules, while the negative ones exploit the fact that the ℓ^1-norm is not strictly convex.

1. Introduction

The number of possible social choice rules is huge, and even the number of those singled out in the literature for analysis is large. Researchers have tried many different axioms in order to classify and characterize these rules, sometimes leading to impossibility theorems. New rules are still being introduced, and the subject is far from tidy.

A promising unifying framework is that of *distance rationalization* (abbreviated DR), whereby some subset D (the *consensus set*) of the set of elections is distinguished. This subset is further partitioned into finitely many subsets, on each of which there is a different social outcome. We choose a distance measure d on elections, and for each election E outside D, an outcome is chosen socially if and only if it is chosen in some profile in D minimizing the distance under d to E. The basic idea is very old (it is essentially a generalization of the maximum likelihood interpretation of voting stemming from work of Condorcet), and has been studied in particular by Nitzan and coauthors [18, 1]. This approach "decomposes" a rule into simpler components D and d.

More recently, distance rationalization was explicitly formalized as a unifying principle by Meskanen and Nurmi [15]. Elkind, Faliszewski, and Slinko have further developed the theory [9, 10, 7, 8, 6]. In particular, they have studied sufficient conditions on consensus and distance for the resulting rule to satisfy desirable properties such as monotonicity, anonymity, homogeneity. Their work shows that almost every rule can be represented in the DR framework, and the main interest in the subject is when we can choose the distance and consensus notations to be natural and computationally tractable.

Our motivation here is to explore in detail the relationship between the distance rationalizability framework and other representations of voting rules. In particular, we wish to specialize to the case of anonymous and homogeneous rules, which include most rules used in practice, and most appearing in the research literature. In this case there is a known geometric interpretation using simplices, popularized by Saari [20], which is very useful for considering

Key words and phrases. social choice theory, collective decision-making, rankings, homogeneity, anonymity, simplex, Wasserstein, Earth Mover distance.

We thank John Hillas for raising the question leading us to Example 2.26 and Proposition 2.27. The second author thanks Elchanan Mossel and Miklos Racz for useful conversations.

many questions. We aim to clarify just how the geometric idea of distance rationalization connects to the simplex representation, and use this to sharpen results about anonymous and homogeneous rules using the distance rationalizability framework.

1.1. **Our contribution.** Our approach, of deriving general results about distance rationalizability, rather than special results for a few known rules, is similar to that of Elkind, Faliszewski and Slinko, whose results we improve in some places.

We deal with a large number of foundational and definitional issues, many of which have not been discussed by previous authors (in some cases because the level of generality they operated in was not sufficient to distinguish these concepts). These questions must be considered systematically in order for us to make further progress. Some of our contribution consists of a more efficient and rigorous presentation of known material. However, most of the present article deals with new ideas.

Section 2 develops the basic notation and terminology. Our approach is similar to that taken by previous researchers, but there are some improvements. We aim to operate generally (for example, by using a hemimetric rather than a metric), and explicitly distinguish several concepts that have sometimes been conflated in previous work. We pose, and give a partial answer to in Example 2.26 and Proposition 2.27, an important question regarding uniqueness of representation in the DR framework, which does not appear to have been noticed before.

In Section 3 we deal with anonymous rules. We characterize anonymous DR rules in Proposition 3.8. We go deeper in the case of *standard* distances, and obtain Proposition 3.16 which connects the two definitions of distance rationalizability. The definition of totally anonymous distance and the formula for the quotient distance in Proposition 3.14 are important, and, as far as we can tell, new. This section shows the connection between ℓ^1-votewise distances and the Earth Mover distance.

All previous work on distance rationalizability of which we are aware has dealt directly with profiles, and hence has a discrete flavour. Section 4 goes further, to connect the DR representation with the usual simplex representation, allowing an interpretation in terms of continuous mathematics. This leads to Proposition 4.5, which looks obvious only because the definitions have been correctly made. We focus in Section 4.2 on the special case of ℓ^p-votewise distances, where the connection to Wasserstein distances, Minkowski spaces, and the continuous theory of optimal transportation is very clear. This gives a completely new perspective not yet seen in the voting literature, and suggests not only new voting rules but also a possible way to attack harder questions in the DR framework. A weakness of the DR framework is that good sufficient conditions for a rule to be homogeneous are not known. We make some modest improvements in Proposition 4.10.

In Section 6 we explore the decisiveness of homogeneous DR rules. On the positive side, we give a sufficient condition in Corollary 6.5 for rules from a wide class to be *hyperplane rules* and therefore be rather decisive. In particular, any rule defined using an ℓ^p votewise norm and the strong or weak unanimity consensus satisfies this condition. On the negative side, we show in Proposition 6.9 that large tied sets occur often with ℓ^1 votewise metrics, and subtle interactions between consensus and distance must be analysed. We conclude in Section 7 with some ideas for future work.

2. Basic definitions

We use the standard conventions of social choice theory.

Definition 2.1. *Let C be a finite set of m* **candidates** *and V a finite set of n* **voters***. For each s with $1 \leq s \leq m$, an s-***ranking** *is a strict linear order of s elements chosen from C. The set of all s-rankings is denoted $L_s(C)$. When $s = m$ we write simply $L(C)$. The set of* **profiles** *is $\mathcal{P} := \mathcal{P}(C, V) := L(C)^V$. An* **election** *is a triple (C, V, π) with $\pi \in \mathcal{P}(C, V)$. We denote the set of all elections by $\mathcal{E}$.*

A **social rule of size** *s is a function that takes each election to a subset of $L_s(C)$. When there is a unique s-ranking chosen, the word "rule" becomes "function". When $s = 1$, we have the usual* **social choice function**, *and when $s = m$ the usual* **social welfare function**.

For each subset D of $\mathcal{E}$ (a **domain restriction**) we can consider a D**-social rule** to be defined as above, but with domain restricted to D.

2.1. **Consensus.** A consensus is simply a socially agreed unique outcome on some set of elections.

Definition 2.2. *Let $1 \leq s \leq m$. An s-***consensus** *is a D-social function of size s, for some $D \neq \emptyset$. We usually denote a consensus by $\mathcal{K}$. The domain D of $\mathcal{K}$ is called an s-***consensus set** *and is partitioned into the inverse images $\mathcal{K}_r := \mathcal{K}^{-1}(\{r\})$, with $r \in L_s(C)$. We will sometimes identify $\mathcal{K}$ with its domain $D(\mathcal{K}) = \cup_{r \in L_s(C)} \mathcal{K}_r$. If $\mathcal{K}, \mathcal{K}'$ are s-consensuses with $\mathcal{K}_r \subseteq \mathcal{K}'_r$ for each $r \in L_s(C)$, we say that $\mathcal{K}$ is a* **refinement** *of $\mathcal{K}'$ and write $\mathcal{K} \subseteq \mathcal{K}'$.*

The integer s is part of the definition. Every s-consensus $\mathcal{K}$ yields an s'-consensus for each $s' \leq s$, by restricting to an initial segment of the preference order. We call this the s'**-restriction** of $\mathcal{K}$.

Several specific consensuses have been described in the literature. We unify the presentation here.

Definition 2.3. *(unanimity consensus)*

Let $1/2 \leq \alpha < 1$. The (α, s)**-unanimity consensus** $\mathbf{S}^{(\alpha,s)}$ *is the s-consensus with domain consisting of all profiles in which there is some fraction strictly more than α of voters, all of whom agree on the order of the top s candidates.*

The limiting value as $\alpha \to 1$ is the case of complete unanimity. We denote this by $\mathbf{S}^s$, the **strong unanimity consensus**. *When $s = m$, we simply write $\mathbf{S}$, whereas when $s = 1$, we denote it $\mathbf{W}$, the* **weak unanimity consensus**.

When $\alpha = 1/2$, we obtain the **majority s-consensus** $\mathbf{M}^s$.

Definition 2.4. *(Condorcet consensus)*

*Let $1/2 \leq \alpha \leq 1$. The (α, s)-***Condorcet consensus** $\mathbf{C}^{(\alpha,s)}$ *has domain consisting of all elections for which an (α, s)-Condorcet partial ranking exists. That is, there is a chain $c_0, \dots, c_{s-1}$ such that for each i and each $c \in C \setminus \{c_0, \dots, c_i\}$, a fraction strictly greater than α of voters rank c_i above c. When $s = 1$ we obtain the concept of* α**-Condorcet winner**; *such a candidate is necessarily unique.*

When $\alpha = 1/2$ we denote this by $\mathbf{C}^s$, and when furthermore $s = 1$, we write simply $\mathbf{C}$. When $\alpha \to 1$, we obtain $\mathbf{S}^s$.

Definition 2.5. *("single-peaked consensus")*

Consider the set of **single-peaked** *profiles (those for which there is a fixed ordering of C with respect to which the following is true: for each voter v, there is an "ideal" element c_i such that if $k \leq j \leq i$ or $i \leq j \leq k$, v prefers c_i to c_j to c_k).*

If $n := |V|$ is odd, the median of the ideal elements c_i is the consensus winner. In this case, for each $r \in L_1(C)$, every single-peaked election with median element r also belongs to $\mathbf{C}_r$. When n is even, it is not clear how to define **SP**.

Definition 2.6. *The* **positional scoring rule** *defined by a vector w with $1 = w_1 \geq \cdots \geq w_m = 0$ elects all candidates with maximal score, where the score of a in the profile π is defined as $\sum_{v \in V} w_{\pi(v)^{-1}(a)}$. Consider the set of all profiles for which all positional scoring rules yield the same unique winner. By convexity of the set of weight vectors, it suffices to check the finite set of rules defined by weight vectors $(1, 1, \dots, 1, 0, \dots 0)$ where the number of 1's is k; in other words the* k**-approval rules** *for $1 \leq k \leq m - 1$. We denote this consensus $\mathbf{\Sigma}$. It has been called the* **Lorenz consensus**.

Remark 2.7. $\mathbf{\Sigma}$ *contains $\mathbf{W}$ but otherwise there is no containment relation between $\mathbf{\Sigma}$ and any of the other consensuses above. Another interpretation is that a is the winner in $\mathbf{\Sigma}$ if and only if a has the most first-place votes, the greatest total of first- and second-place votes, etc.*

Each of the consensuses described above is a social rule satisfying *anonymity*, meaning that the consensus is invariant under any permutation of the voter set. Each also satisfies several other normative properties which we discuss in Section 5. The basic idea behind distance rationalizability is to extend each consensus in a simple way to a social rule with full domain, while keeping as many of these properties as possible. Usually, it is not possible to retain all desirable properties (social choice theory is notorious for such impossibility results). Thus consensuses are strongly related to the topic of *domain restrictions*.

2.2. **Distances.** We require a notion of distance on elections. We aim to be as general as possible.

Definition 2.8. *(distance) By a* **distance** *(or* **hemimetric***) on $\mathcal{E}$ we mean a function $d : \mathcal{E} \times \mathcal{E} \to \mathbb{R}_+ \cup \{\infty\}$ that satisfies the identities $d(x, x) = 0$ and $d(x, z) \leq d(x, y) + d(y, z)$. A distance that is symmetric ($d(x, y) = d(y, x)$) and distinguishes points ($d(x, y) = 0 \Rightarrow x = y$) is called a* **metric***. We call a distance* **standard** *if $d(E, E') = \infty$ whenever E and E' have different sets of voters or candidates.*

One commonly used class of distances consists of the **votewise** distances formalized in [10]. They are each based on distances on $L(C)$. See [5] for basic information about metrics on the symmetric group.

Example 2.9. *The most commonly used such distances on $L(C)$ are:*

- *the* **discrete metric** *d_H, with $d_H(\rho, \sigma) = 1$ if and only if $\rho \neq \sigma$, and $d_H(\rho, \rho) = 0$;*
- *the* **inversion metric** *d_K (also called the swap, bubblesort or Kendall-τ metric), where $d_K(\rho, \sigma)$ is the minimal number of swaps of adjacent elements required to convert ρ to σ;*
- **Spearman's footrule** *d_S, defined to be $\sum_i |\rho(i) - \sigma(i)|$, where we interpret $L(C)$ as permutations of $\{1, \ldots, n\}$.*

Recall that a *seminorm* is a real-valued function N on $\mathbb{R}^n$ such that $N(x+y) \leq N(x)+N(y)$ and $N(\lambda x) = |\lambda|N(x)$ for all $\lambda \in \mathbb{R}$ and all $x, y \in \mathbb{R}^n$. It follows that $N(x) \geq 0$. If $N(x) = 0$ only if $x = 0$, then we say N is a **norm**. A commonly used norm is the ℓ^p-norm given by $||x||_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for fixed $1 \leq p < \infty$. The ℓ^∞ norm $||x||_\infty = \max_{1 \leq i \leq n} |x_i|$ has also been used.

Definition 2.10. *(votewise distances)*

Fix a distance d on $L(C)$, and choose a family $\{N_n\}_{n \geq 1}$ of seminorms, where N_n is defined on $\mathbb{R}^n$. Extend d to a function on $\mathcal{P}$ by taking $n = |V|$ and defining for $\sigma, \pi \in \mathcal{P}$

$$d_N(\pi, \sigma) := N_n(d(\pi_1, \sigma_1), \ldots, d(\pi_n, \sigma_n)).$$

This yields a distance on elections having the same profile set. We complete the definition of the extended distance (which we denote by d_N) on $\mathcal{E}$ by declaring it to be standard.

We use the abbreviation d^p for d_{ℓ^p}, and sometimes we even use just d for d_N if the meaning is clear.

Note that if d is a metric and N is a norm, then d_N is a metric.

Example 2.11. *(famous votewise distances)*

The ℓ^1-votewise distances on $\mathcal{E}$ based on d_H and d_K are called respectively the **Hamming metric** *and* **Kemeny metric***. To avoid confusion, note that some authors have considered the Hamming metric on $L(C)$ itself. The Hamming metric measures the number of voters whose preferences must be changed in order to convert one profile to another, and as such has an interpretation in terms of bribery. The Kemeny metric measures how many swaps of adjacent candidates in a single preference order are required, and is related to models of voter error.*

Among the many other votewise metrics, we single out the ℓ^1-votewise metric based on d_S, sometimes called the **Litvak distance** *d_L.*

Example 2.12. *(tournament distances) Given an election E, we form the pairwise tournament digraph with nodes indexed by the candidates, where the arc from a to b has weight equal to the net support for a over b in a pairwise contest. Note that there are some redundant arcs compared to the definitions of some other authors.*

*The weighted adjacency matrix $M(E)$ is antisymmetric. Given a matrix seminorm N, we define the N-***tournament distance** *by $d_N(E, E') = N(M(E) - M(E'))$. A closely related distance is defined in the analogous way, but where each element of the adjacency matrix is replaced by its sign (1, 0, or -1). We call this the N-***reduced tournament distance***. We denote the special cases where N is the ℓ^1 norm on matrices by d_T and d_{RT} respectively. A (reduced) tournament distance cannot be a metric, even if N is a norm, because it does not distinguish points — small profile changes usually do not change the majority tournament. However, it is symmetric (a* **pseudometric***).*

Example 2.13. *(nonsymmetric and nonstandard distances) In this paper we focus on symmetric standard distances. Nonsymmetric distances occur when considering the cost of changing a vote (whether through bribery or for reasons known only to the voter). For example, it may be much more costly (for social reasons) to change a vote away from unanimity than towards it. A distance that fails to be a metric only because of asymmetry is called a* **quasimetric***.*

Some rules, for example Young's rule, are defined in terms of deletion of voters and for this nonstandard distances are needed. For example, let $d_{del}(E, E')$ (respectively $d_{ins}(E, E')$) be defined as the minimum number of voters we can delete from (insert into) election E in order to reach election E' (or $+\infty$ if E' can never be reached). Each is nonstandard, and so are their symmetrized versions, which are metrics [9].

Example 2.14. *(shortest path distances)*

We can use $\mathcal{E}$ as the underlying set of a digraph, by defining an arc between E and E' if and only if $d(E, E') = 1$. Then the length of a shortest path on such a digraph gives a quasimetric d', which is a metric if the underlying digraph is a graph. Sometimes, $d' = d$. For example, $d_H, d_K, d_{ins}, d_{del}$ are essentially defined via this construction.

While shortest path distances occur often in the literature, they are rather special.

Proposition 2.15. *A quasimetric on $\mathcal{E}$ is a shortest path distance for some nonempty edge relation on $\mathcal{E}$ if and only if it takes integer values, and for each $y, x \in \mathcal{E}$ such that $2 \leq d(y, x) < \infty$, there is $z \in \mathcal{E}, z \neq x, z \neq y$ such that $d(y, z) + d(z, x) = d(y, x)$.*

Proof. Each shortest path quasimetric satisfies the given conditions. For the converse, suppose that d is a quasimetric on $\mathcal{E}$ satisfying the given conditions. It follows that for every y, x, there are $x_0, \dots, x_k$ such that $d(y, x) = k$ and each $d(x_i, x_{i+1}) = 1$. Define an arc between E and E' if and only if $d(E, E') = 1$. Let d' be the shortest path distance on the associated digraph. It follows by induction on the minimal value of k that $d = d'$. □

Note that if we allowed arbitrary nonnegative weights on the edges of the digraph, every distance could be represented as a shortest path distance.

2.3. **Combining consensus and distance.** In order for a rule to be definable via the DR construction, it is necessary that the following property holds, and we shall assume this from now on.

Definition 2.16. *Let d be a distance and $\mathcal{K}$ a consensus. Say that $(\mathcal{K}, d)$* **distinguishes consensus choices** *if whenever $x \in \mathcal{K}_r, y \in \mathcal{K}_{r'}$ and $d(x, y) = 0$, then $r = r'$.*

We use a distance to extend a consensus to a social rule in the natural way. The choice at a given election E consists of all s-rankings r whose consensus set $\mathcal{K}_r$ minimizes the distance to E. We introduce the idea of a score in order to use our intuition about positional scoring rules. Note that if $\mathcal{K}_r$ is empty, then $|r| = \infty$.

$\mathcal{K}/d$	$\mathbf{S}$	$\mathbf{W}$	$\mathbf{C}$	$\mathbf{C}^m$
d_K^1	Kemeny	Borda	Dodgson	
d_H^1	modal ranking	plurality	VRR	
d_L	Litvak	Borda	Dodgson	
d_T	Kemeny	Borda	maximin	
d_{RT}	Copeland	Copeland	Copeland	Slater
d_{ins}	trivial	trivial	maximin	
d_{del}	modal ranking	plurality	Young	

TABLE 1. Some known rules in the DR framework

Definition 2.17. *Let $1 \leq s \leq m$ and suppose that $\mathcal{K}$ is an s-consensus and d a distance on $\mathcal{E}$. Fix an election $E \in \mathcal{E}$. The* $(\mathcal{K}, d, E)$**-score** *of $r \in L_s(C)$ is defined by*

$$|r| := \min_{E' \in \mathcal{K}_r} d(E, E').$$

The rule $R := \mathcal{R}(\mathcal{K}, d)$ is defined by

$$R(E) = \arg\min_r |r|. \tag{1}$$

We say that R is **distance rationalizable** *(DR) with respect to $(\mathcal{K}, d)$.*

Note that the scores are defined so that they are nonnegative, and higher score corresponds to larger distance. This is not consistent with the usual scoring rule interpretation in Example 2.18 — our DR scores typically have the form $M - s$ where s is the score associated with the rule and M depends on E but not on any $r \in L_s(C)$.

Table 1 presents a few known rules in this framework. Most of these entries are well known. We single out the following less obvious references: modal ranking rule [2]; voter replacement rule [9]. The entries marked "trivial" are so labelled because in those cases every profile not in $\mathcal{K}$ is at distance $+\infty$ from every $\mathcal{K}_r$. Missing entries reflect on the authors' knowledge, and may have established names. Our table overlaps with that in [15] — note that the $(\mathbf{C}, d_H^1)$ entry is incorrect in that reference.

Example 2.18. *(scoring rules) Every positional scoring rule defined by weight vector w with $w_1 \geq w_2 \geq \cdots \geq w_m$ and $w_1 > w_m$ has the form $\mathcal{R}(\mathbf{W}, d_w^1)$ where d_w is the distance on rankings defined by*

$$d_w(\rho, \rho') = \sum_{c \in C} |w_{\rho^{-1}(c)} - w_{\rho'^{-1}(c)}|.$$

Plurality and Borda are special cases, where d_w simplifies to d_H^1 and d_L respectively. As far as the distance to $\mathbf{W}$ or $\mathbf{C}$ is concerned, d_L and d_K are proportional, but contrary to the misstatement in [9, Section 3], they are not proportional in general [15, p. 298–299].

Remark 2.19. *Note that d_w is a metric on $L_s(C)$ if and only if $w_1, \ldots, w_s$ are all distinct. The score of r under the rule defined by w is the difference $nw_1 - |r|$. For example, for Borda with m candidates, the maximum possible score of a candidate c is $(m-1)n$, achieved only for those profiles in $\mathbf{W}_c$. The score of c under Borda is exactly $(m-1)n - K$ where K is the total number of swaps of adjacent candidates needed to move c to the top of all preference orders in $\pi(E)$.*

Example 2.20. *(Copeland's rule)* **Copeland's rule** *can be represented as $\mathcal{R}(\mathbf{C}, d_{RT})$. Indeed, in an election E, the Copeland score of a candidate c (the number of points it scores in pairwise contests with other candidates) equals $n - 1 - s$, where s is the minimum number of pairwise results that must be changed for E to change to a profile that belongs to $\mathbf{C}_c$.*

Example 2.21. *(maximin rule)*
The **maximin** *rule is $\mathcal{R}(\mathbf{C}, d_{ins})$ [9]. The DR score is related to the maximin score in a similar way to the previous example.*

Every rule $\mathcal{R}(\mathcal{K}, d)$ where $\mathcal{K}$ is a 1-consensus automatically yields a social rule $\mathcal{R}^s(\mathcal{K}, d)$ as follows.

Definition 2.22. *Let $1 \leq s \leq m$ and suppose that $\mathcal{K}$ is a 1-consensus and d a distance on $\mathcal{E}$. We define a social rule $\mathcal{R}^s(\mathcal{K}, d)$ of size s as follows. For each $(C, V, \pi) \in \mathcal{E}$, report the top s alternatives in descending order of score. If ties occur, report all possible such subsets of s alternatives.*

Remark. *If the s'-consensus $\mathcal{K}'$ is the s'-restriction of the s-consensus $\mathcal{K}$, it is not necessarily the case that $\mathcal{R}(\mathcal{K}, d) = \mathcal{R}^s(\mathcal{K}', d)$. For example, consider the social welfare rule $\mathcal{R}(\mathbf{M}^m, d_H)$. This elects the modal ranking, the ranking represented most often among voters. However $\mathcal{R}^m(\mathbf{M}^1, d_H)$ ranks alternatives in descending order of plurality score.*

2.4. **Existence and uniqueness.** The DR framework is not very restrictive without further assumptions on $\mathcal{K}$ and d. We give necessary and sufficient conditions, a slight extension of [10, Theorem 3.1].

Definition 2.23. *For each rule R, there is a unique* **maximal consensus** *$\mathcal{K}^*$, namely that whose consensus set D^* consists of all elections on which R gives a unique output, which output we define as the consensus choice.*

Proposition 2.24. *A rule R is distance rationalizable if and only if every $r \in L_s(C)$ that is chosen at some profile is also the unique choice at some profile. That is, for each $r \in L_s(C)$, if there exists $E \in \mathcal{E}$ such that $r \in R(E)$, then there exists $E' \in \mathcal{E}$ such that $R(E') = \{r\}$.*

Proof. The condition is necessary, because no r can occur as a winner in $\mathcal{R}(\mathcal{K}, d)$ unless it wins somewhere in $\mathcal{K}$. If the condition holds, take $\mathcal{K} = \mathcal{K}^*$. We can recapture R as $\mathcal{R}(\mathcal{K}, d)$ by defining (as in the proof of [10, Theorem 3.1]) d to be the shortest path distance on the graph having an edge between E and E' if and only if exactly one of E and E' (without loss of generality, E) belongs to $D(\mathcal{K})$, and $R(E) \subset R(E')$. Note that if $r \neq r'$ and $R(E) = \{r\}, R(E') = \{r'\}$, then $d(E, E') \geq 2$, while if $|R(E)| \geq 2$, E is at distance 1 from precisely those $\mathcal{K}_r$ for which $r \in R(E)$. □

Remark 2.25. *If a rule satisfies the usual* **nonimposition** *axiom (every r can win uniquely in some profile), then by Proposition 2.24 it is distance rationalizable [10, Theorem 3.1]. However this assumption is not necessary — consider the rule in which $R(E) = \{a\}$ for every election E, and use an arbitrarily chosen distance.*

Furthermore [9, Theorem 3.3] a rule R can be distance-rationalized by some d with respect to a fixed $\mathcal{K}$ if and only if R agrees with $\mathcal{K}$ on its domain (called *compatibility* in [9]). Thus for example, every rule satisfying the usual unanimity axiom can be rationalized with respect to $\mathbf{S}$. In view of the flexibility of the DR framework, it is clear that the key idea is to make an appropriate choice of a "small" $\mathcal{K}$ and "natural" d so as to recapture rule R via $R = \mathcal{R}(\mathcal{K}, d)$.

We now turn to the question of how much flexibility there is in the choice of $\mathcal{K}$ and d. The construction in the proof of Proposition 2.24 shows that changing both of them can lead to the same rule. When $\mathcal{K}$ is fixed and d varies, the rule often changes, but not always. A general class of examples where the rule does not change is discussed at the end of Section 3.1.

Similarly, when d is fixed and $\mathcal{K}$ varies, the rule sometimes does not change. For example, consider Copeland's rule, which can be described as $\mathcal{R}(\mathbf{C}, d_{RT})$. It can also be described as $\mathcal{R}(\mathbf{W}, d_{RT})$, because for each $a \in C$, every point of $\mathbf{C}_a$ is at distance zero from $\mathbf{W}_a$ with respect to d_{RT}. A key foundational question is: if R has the form $\mathcal{R}(\mathcal{K}, d)$, is it necessarily the case that $R = \mathcal{R}(\mathcal{K}^*, d)$? The answer is no in general, as we now show.

Example 2.26. *Let $\mathcal{K} = \mathbf{S}$ and fix a ranking r. We define d for elections y and z on same voter set and candidate set by*

$$d(y, z) = \begin{cases} 0 & \text{if } y = z \\ d_H^1(y, z) + 42 & \text{if } y \in \mathbf{S}_r \text{ or } z \in \mathbf{S}_r \\ d_H^1(y, z) & \text{if } y \notin \mathbf{S}_r \text{ and } z \notin \mathbf{S}_r. \end{cases}$$

and we complete this definition by setting d to take $+\infty$ if the voter or candidate sets differ. It is easily checked that d is a standard distance. Then, consider elections on 43 voters. All elections which have 42 voters voting r and one voting $r' \neq r$ are tied. But elections where 41 voters vote for r and 2 vote for r' are won by r'. Thus, these elections are in $\mathcal{K}^*$, so $\mathcal{K}^*_{r'}$ is considerably extended compared to $\mathcal{K}_{r'}$. On the other hand, $\mathcal{K}^*_r = \mathcal{K}_r$ is unchanged. Thus, in $\mathcal{R}(\mathcal{K}^*, d)$, elections where 42 voters vote r and one votes $r' \neq r$ are won by r' and are not tied anymore.

However, in many cases, the equality $R = \mathcal{R}(\mathcal{K}^*, d)$ is true. We denote by $\mathcal{K}^\epsilon$ the refinement of $\mathcal{K}^*$ containing all elections at distance at most ϵ under d from $\mathcal{K}$.

Proposition 2.27. *If $R = \mathcal{R}(\mathcal{K}, d)$ and d is a shortest path distance, then $R = \mathcal{R}(\mathcal{K}^\epsilon, d)$ for all ϵ. In particular, $R = \mathcal{R}(\mathcal{K}^*, d)$.*

Proof. Consider an election E. If $E \in \mathcal{K}^\epsilon$ then it has the same winner according to R and $\mathcal{R}(\mathcal{K}^\epsilon, d)$ because $\mathcal{K}$ is a refinement of $\mathcal{K}^\epsilon$. Now suppose that $E \notin \mathcal{K}^\epsilon$ and consider a ranking r. The distance from E to $\mathcal{K}_r$ is finite if and only if there exists a path in the underlying graph of d from E to $\mathcal{K}_r$. Then, by definition of $\mathcal{K}^\epsilon$, the last $\lfloor \epsilon \rfloor$ points of this path are in $\mathcal{K}^\epsilon$, and all the other points are not. So the distance from E to $\mathcal{K}^\epsilon_r$ is at most $d(E, \mathcal{K}_r) - \lfloor \epsilon \rfloor$. Conversely, $d(E, \mathcal{K}_r) - \lfloor \epsilon \rfloor \geq d(E, \mathcal{K}^\epsilon_r)$, because, one can extend a path from E to $\mathcal{K}^\epsilon_r$ in a path from E to $\mathcal{K}_r$.

Now, for any $E \in \mathcal{K}^*$, there exists ϵ such that $E \in \mathcal{K}^\epsilon$. Then, since $\mathcal{K}^\epsilon$ is a refinement of $\mathcal{K}^*$, we have that E has the same winner according to R and to $\mathcal{R}(\mathcal{K}^*, d)$. □

3. Anonymity

Anonymous social rules occur very often in practice. In this section, we show how to express distance rationalization for anonymous rules using only anonymous objects and functions.

Informally, a voting rule is **anonymous** if it depends only on the set of votes cast, not on the identity of the voters. Most commonly used rules have this property. In order to define an anonymous rule in the DR framework, we need to discuss anonymous distances and consensuses. The second concept is obvious, but the first contains some subtleties. We shall therefore be more formal than usual.

Example 3.1. *Consider $R := \mathcal{R}(\mathbf{W}, d^1_H)$, plurality rule. The consensus $\mathbf{W}$ is anonymous, because identities of voters are not important (they all vote the same way in the consensus). The distance d^1_H is* not *anonymous, however, in the following sense. Consider the case of two voters and three candidates, and the profiles $\pi = (abc, acb), \sigma = (acb, abc)$. Then $d(\pi, \sigma) \neq 0 = d(\pi, \pi)$, although both π, σ belong to $\mathbf{W}_a$ and one is a permutation of the other.*

Note that below we do use the term "anonymous" in a way that includes d^1_H, and "totally anonymous" to describe the stronger property above.

One obvious solution to this problem is to consider, instead of d itself, the quantity $\min_{e,e'} d(e, e')$ where the minimum is taken over all e equivalent to E and e' equivalent to E'. This construction is intuitive when we think of d as measuring cost of changing profiles: since the ordering of voters should be irrelevant, we simply choose the optimal ordering of voters to minimize the cost. We make this idea precise below.

3.1. Formal description.

We can define two elections E, E' to be equivalent if they have the same candidate sets and the multiset of votes associated to each election is the same. A rule is anonymous if and only if it respects this equivalence relation. In other words, the identities of the voters are irrelevant, as is any ordering of them. We give an equivalent, slightly different, and more formal definition below.

Definition 3.2. *Define the* **vote number map** *$\mathcal{N}$ on $\mathcal{E}$ by sending $E = (C, V, \pi)$ to the map that assigns to each $\rho \in L(C)$ the number of entries of π equal to ρ:*

$$\mathcal{N}(E)_\rho = \#\{v \mid \pi_v = \rho\}.$$

We denote by $\mathcal{V}$ the quotient of $\mathcal{E}$ modulo $\mathcal{N}$, and the quotient map by $E \mapsto \overline{E}$.

By a **partial social rule on $\mathcal{V}$ of size** *s we mean a function defined on a subset of $\mathcal{V}$, taking values in $L_s(C)$.*

Definition 3.3. *Let R be a partial social rule. Then R is* **anonymous** *if and only if R is constant on each fibre of $\mathcal{N}$.*

Proposition 3.4. *The following conditions are equivalent for a partial social rule R of size s.*

(i) R is anonymous.
(ii) There is a partial social rule $\overline{R}$ of size s on $\mathcal{V}$, such that $R(E) = \overline{R}(\overline{E})$ for every $E \in \mathcal{E}$.

□

Remark 3.5. *For each $E = (C, V, \pi)$, define $G(E)$ to be the group of all bijections of V. For each $E \in \mathcal{E}$ and each $g \in G(E)$, define $g(E) = (C, V, g(\pi))$, where $g(\pi)$ is the profile for which voter $g(v)$ votes as v did in π. If R is anonymous, then $R(g(E)) = R(E)$ for all $E \in \mathcal{E}$ and all $g \in G(E)$. Other authors take this as the definition of anonymity.*

From now on we aim to describe $\overline{R}$ in the DR framework without any explicit mention of profiles, but only using $\mathcal{V}$. We can define a rule to be **distance-rationalizable in $\mathcal{V}$** if it can be defined via a consensus and distance on $\mathcal{V}$ analogously to Definition 2.17. The main question is how such rules relate to rules defined on $\mathcal{E}$. We need to discuss anonymity of distance in some detail.

Definition 3.6. *A distance is* **totally anonymous** *if $d(E, E') = d(F, F')$ whenever $\mathcal{N}(E) = \mathcal{N}(F)$ and $\mathcal{N}(E') = \mathcal{N}(F')$.*

Remark 3.7. *In particular $d(g(E), g'(E')) = d(E, E')$ for all $E, E' \in \mathcal{E}$ and all $g \in G(E), g' \in G(E')$. Note that we do not require d to be standard. Note that a totally anonymous distance d is not a metric, because whenever $\mathcal{N}(E) = \mathcal{N}(E')$, necessarily $d(E, E') = 0$.*

Proposition 3.8. *The following conditions are equivalent for a social rule R.*

(i) R is anonymous and distance rationalizable.
(ii) $R = \mathcal{R}(\mathcal{K}, d)$ for some $\mathcal{K}$ and d, where $\mathcal{K}$ is anonymous and d is totally anonymous.
(iii) $\overline{R}$ is distance-rationalizable in $\mathcal{V}$, $\overline{R} = \mathcal{R}(K, \delta)$ for some K and δ.

If these conditions hold, then we can choose d and δ so that $\delta(\overline{E}, \overline{E'}) = d(E, E')$ for all $E, E' \in \mathcal{E}$.

Proof. To show that the first condition implies the second, the construction of Proposition 2.24 suffices. Now suppose that $R = \mathcal{R}(\mathcal{K}, d)$ for some $\mathcal{K}$ and d, where $\mathcal{K}$ is anonymous and d is totally anonymous. Then R is constant on each fibre of $\mathcal{N}$ and $\overline{R} = \mathcal{R}(\overline{\mathcal{K}}, \delta)$ where δ is defined by $\delta(\overline{E}, \overline{E'}) = d(E, E')$. Note that δ is well defined precisely because d is totally anonymous. Finally, if $\overline{R} = \mathcal{R}(K, \delta)$ then clearly R is anonymous and by composing with the map $\mathcal{N}$, we can define a consensus $\mathcal{K}$ and distance d on $\mathcal{E}$, such that $R = (\mathcal{K}, d)$ satisfies $\overline{R} = S$, $\overline{\mathcal{K}} = K$ and $d(E, E') = \delta(\overline{E}, \overline{E'})$ for all $E, E' \in \mathcal{E}$. Then d is totally anonymous. □

Totally anonymous distances occur often in practice.

Proposition 3.9. *The following results hold.*

(i) Every tournament distance is totally anonymous.
(ii) No votewise distance is totally anonymous.
(iii) d_{ins} and d_{del} are totally anonymous.

□

Thus rules based on tournament distances and anonymous consensus, such as Copeland's rule or the minimax rule, relate in an obvious way to rules defined on $\mathcal{V}$, and distance computations in $\mathcal{V}$ correspond directly to distance computations in $\mathcal{E}$. However, it turns out that for rules defined by distances that are not totally anonymous, such as votewise distances, the situation is more complicated. We know that such a rule can also be defined by a totally anonymous distance, but we want to do this while not changing the consensus. This is possible in some cases, as we now proceed to show.

Definition 3.10. *A standard distance d on $\mathcal{E}$ is* **anonymous** *if it satisfies $d(g(E), g(E')) = d(E, E')$ for all $E, E' \in \mathcal{E}$ such that $E = (C, V, \pi)$ and $E' = (C, V, \pi')$, and all $g \in G(E)$.*

Clearly, every totally anonymous distance is anonymous, but the converse is not true.

Example 3.11. *A votewise distance is anonymous if each of the family of norms defining it is symmetric (meaning that it is invariant under permutations of its arguments). In that case, letting $s = \{s_1^{n_1}, \dots, s_k^{n_k}\}$ denote the multiset of weight n corresponding to the n-tuple $(a_1, \dots, a_n) \in \mathbb{R}^n$, we can define $N(s) = N_n(a)$, where $n = \sum_i n_i s_i$ can be computed knowing only s.*

The following is a generalization of [10, Thm 4.5], which was stated for votewise distances and four specific consensuses.

Proposition 3.12. *If $\mathcal{K}$ is anonymous and d is standard and anonymous, then $\mathcal{R}(\mathcal{K}, d)$ is anonymous.*

Proof. Let $E = (C, V, \pi)$ and $E' = g(E) = (C, V, \pi')$ be two elections with the same candidate and voter sets. Since $\mathcal{K}$ is anonymous, $g(\mathcal{K}_r) = \mathcal{K}_r$ for all g and r. Then $d(E, \mathcal{K}_r) = d(g(E), g(\mathcal{K}_r)) = d(E', \mathcal{K}_r)$. □

When d is anonymous and standard, there is a simple formula which allows us to derive a similar result to Proposition 3.8.

Definition 3.13. *The general definition of quotient distance with respect to an equivalence relation $\sim$ uses a shortest path construction:*

$$\overline{d}(x, y) = \min d(E, E_1') + d(E_2, E_2') + \dots + d(E_k, E') \tag{2}$$

where the minimum is taken over all admissible paths*, namely paths such that $E_i' \sim E_{i+1}$ for $1 \le i \le k-1$, E projects to x and E' to y.*

We apply this in the case where $\sim$ is the kernel of the map $\mathcal{N}$, yielding an induced distance $\overline{d}$ on $\mathcal{V}$.

We now show how to compute $\overline{d}$ from d when d is anonymous. As noted in the introduction to Section 3, each anonymous distance d on $\mathcal{E}$ induces a map on $\mathcal{V}$, via $(x, y) \mapsto \min_{E,E'} d(E, E')$, where E is a preimage under of x and E' is a preimage of y. This type of construction for a general equivalence relation does not satisfy the triangle equality. However, in our main situation of interest, the simpler formula does indeed suffice to define the quotient mapping.

Proposition 3.14. *Let d be an anonymous standard distance on $\mathcal{E}$. Then the quotient distance induced on $\mathcal{V}$ is given by*

$$\overline{d}(x, y) = \min_{E'} d(E, E') = \min_{g} d(E, g(E')) \tag{3}$$

where E is any preimage of x, the first minimum is taken over all preimages E' of y, and the second minimum can be taken over all g for a fixed preimage E' of y.

Proof. We first show that the minimum value of k for paths achieving the minimum in (2) is 0. Assume that this is not the case, so there exist a minimal $k > 0$ and admissible paths such that

$$\overline{d}(x, y) = \sum_{i=1}^{k} d(E_i, E_i') < \infty.$$

Choose g so that $g(E_k) = E'_{k-1}$ (possible by definition of $\sim$). Note that $d(E_k, E'_k) < \infty$ and so since d is standard, E_k and E'_k have the same number of voters. Thus g can be applied to E'_k also. Then by anonymity and the triangle inequality

$$\begin{aligned} d(E_{k-1}, E'_{k-1}) + d(E_k, E'_k) &= d(E_{k-1}, E'_{k-1}) + d(E'_{k-1}, g(E'_k)) \\ &\geq d(E_{k-1}, g(E'_k)). \end{aligned}$$

This contradicts the minimality of k, and this gives the first displayed equality. The second follows because all E' projecting to y are equivalent, so that each can map onto any other via some g. □

Remark 3.15. *Non-standard distances can violate this result. For example, take $d = d'_{ins}$. Then $d(E, E')$ is finite iff $V \subset V'$ or the inverse. So the expression for $\overline{d}$ in (3) is always infinite, because distinct elections E and E' of the same size are at infinite distance. But the quotient metric is finite (use a path going through an election including both E and E').*

Proposition 3.16. *Let d be an anonymous standard distance and $\mathcal{K}$ an anonymous consensus. The anonymous rule $\mathcal{R}(\mathcal{K}, d)$ satisfies $\overline{\mathcal{R}(\mathcal{K}, d)} = \mathcal{R}(\overline{\mathcal{K}}, \overline{d})$, where $\overline{d}$ is as in Proposition 3.14. Furthermore, for every r and E, $\overline{d}(\overline{E}, \overline{\mathcal{K}}_r) = d(E, \mathcal{K}_r)$.*

Proof. We have

$$\overline{d}(\overline{E}, \overline{\mathcal{K}}_r) = \min_{E' \in \mathcal{K}_r} \overline{d}(\overline{E}, \overline{E'}) = \min_{E' \in \mathcal{K}_r} \min_{E'' \sim E'} d(E, E'') = \min_{E'' \in \mathcal{K}_r} d(E, E'') = d(E, \mathcal{K}_r).$$

□

At first sight, Propositions 3.8 and 3.16 may seem inconsistent. The explanation is that given an anonymous standard d and anonymous $\mathcal{K}$, R can be written not only in the form $(\mathcal{K}, d)$ but also in the form $(\mathcal{K}, d')$ where d' is totally anonymous and $\overline{d'} = \overline{d}$. In fact $d'(E, E')$ equals the familiar quantity $\min_g d(E, g(E'))$.

3.2. **Interpretation in terms of optimal transportation.** The problem of optimal transportation originated with Monge in the 19th century and was generalized by Kantorovich in the 1940s. Indeed, such transportation problems were at the heart of Kantorovich's seminal work in linear programming. In the discrete case, the problem amounts to minimizing the cost of transferring mass from one histogram to another while incurring the minimal cost, and can be solved by linear programming.

In the anonymous and standard case, the distance $\overline{d}$ is the solution of an optimal transportation problem, because we must move voter mass between types of voters while incurring the minimal cost (distance). The assumption that d is standard is equivalent to saying that conservation of mass (number of voters) holds. This is a special case of the transportation problem known as the *assignment problem.* The minimum can be computed in polynomial time via the "Hungarian method" [17]. We shall explore this topic more in Section 4.

Example 3.17. *Consider the Hamming distance $d := d^1_H$. For each $x, y \in \mathcal{V}$, the distance $\overline{d}(x, y)$ is the minimum number of voters whose votes must be changed in order to transform x into y. For example if x is a voting situation with 2 abc voters and 3 bac voters, while y has 2 bac voters and 3 cba voters, then $\overline{d^1_H}(x, y) = 3$. Note that for the Kemeny metric $d := d^1_K$, $\overline{d}(x, y) = 8$.*

4. Homogeneity

So far nothing in our definitions has required any coherence in definitions of a voting rule for different values of n. Most voting rules used in practice satisfy the coherence property called homogeneity, which we now explore. From now on we deal only with anonymous rules. Also, we fix C and hence m, and define $M = m!$. In this case there is a well known representation of equivalence classes of elections in terms of the M-dimensional standard simplex Δ. We want to define everything in the DR framework directly on Δ, in terms of fractions of voters with a given preference.

Definition 4.1. *For elections $E = (C, V_1, \pi_1)$ and $E' = (C, V_2, \pi_2)$ where $V_1 \cap V_2 = \emptyset$, we define $E + E'$ to be the election with candidate set C, voter set $V_1 \cup V_2$ and profile π given by $\pi_v = (\pi_1)_v$ if $v \in V_1$ and $(\pi_2)_v$ otherwise. For each $k \geq 1$, define kE to be the election formed by cloning each voter $k - 1$ times and adding repeatedly as above (in an arbitrary but fixed order). An anonymous rule R is* **homogeneous** *if for each E and k, $R(kE) = R(E)$. For a set S of elections, kS means the set of all kE as E ranges over S.*

Let $\Delta(L(C))$ be the set of probability distributions over $L(C)$, and let $\Delta_{\mathbb{Q}}$ consist of all such distributions where the probability of every subset of $L(C)$ is rational. Define the **vote distribution** map that takes each election $E = (C, V, \pi)$ to $x \in \Delta_{\mathbb{Q}}$ by declaring the probability of $\rho \in L(C)$ to be $\mathcal{N}(E)_\rho/|V|$.

An anonymous rule is homogeneous if and only if it is well-defined on $\Delta_{\mathbb{Q}}$. We declare elections E, E' to be equivalent if $E' = kE$ for some $k \geq 1$, and denote the map $\mathcal{E} \mapsto \Delta_{\mathbb{Q}}$ described above by $E \mapsto x := \overline{E}$. From now on we shall deal with homogeneous rules, so there should be no confusion with the related map with the same notation from Section 3. We can now choose a point x of $\Delta_{\mathbb{Q}}$, and unambiguously talk about the winner of the rule at such a point, knowing only the fractions of voters of each type, being free to choose an arbitrary n (provided nx_ρ is an integer for all $\rho \in L(C)$) and preimage election E on n voters. Equivalently, we can simply define a **rule on** $\Delta_{\mathbb{Q}}$ in the obvious way, and this immediately yields an anonymous and homogeneous rule on $\mathcal{E}$.

Adopting this point of view, we can also deal with consensus and distance rationalization. A **consensus in** $\Delta_{\mathbb{Q}}$ is simply a partial social function with domain a subset of $\Delta_{\mathbb{Q}}$. Given a consensus $\mathcal{K}$ and a distance δ on $\Delta_{\mathbb{Q}}$, the rule $\mathcal{R}(\mathcal{K}, \delta)$ is defined in the obvious way.

We now want to connect the two definitions of distance rationalizability.

Definition 4.2. *An anonymous distance on $\mathcal{E}$ is* **homogeneous** *if for each $k \geq 1$ and each $E, E' \in \mathcal{E}$, $d(E, E') = d(kE, kE')$. A family of symmetric seminorms N is* **homogeneous** *if $N(x^{(k)}) = N(x)$ for all $x \in \mathbb{R}^n$ and all $k \geq 1$. Here $x^{(k)}$ denotes the element of $\mathbb{R}^{nk}$ obtained by concatenating k copies of x.*

The reader should avoid confusion by noting that the term *homogeneous* is often used for the different property of a seminorm expressed by the identity $N(\lambda x) = |\lambda| N(x)$.

Example 4.3. *Consider the ℓ^p norm for $1 \leq p \leq \infty$. If we normalize this, the resulting family of norms is homogeneous. Specifically, on $\mathbb{R}^n$ we define $||x||_p^* := n^{-1}||x||_p$.*

Note that being rationalizable with respect to this normalized form is equivalent to being rationalizable with respect to the original norm, since the two norms differ only by a constant multiple.

Proposition 4.4. *Each of the following distances is anonymous and homogeneous.*

- *Every votewise distance based on a homogeneous seminorm.*
- *Every reduced tournament distance.*
- *Every tournament distance based on a homogeneous seminorm.*

□

The proof of the next result is analogous to that of Proposition 3.14.

Proposition 4.5. *Let d be an anonymous and homogeneous standard distance on $\mathcal{E}$. The distance $\overline{d}$ induced by d on $\Delta_{\mathbb{Q}}$ is given by*

$$\overline{d}(x, y) = \min_{E'} d(E, E')$$

where E is of a size n so that there are elections of size n projecting to x and y, $\overline{E} = x$ and the minimum is taken over elections E' of size n such that $\overline{E'} = y$. □

The connection between the two definitions is by now unsurprising.

Proposition 4.6. *If $\mathcal{K}$ and d are anonymous and homogeneous, d is standard, and $\mathcal{R}(\mathcal{K}, d)$ is homogeneous, then $\overline{\mathcal{R}(\mathcal{K}, d)} = \mathcal{R}(\overline{\mathcal{K}}, \overline{d})$.* □

4.1. **Homogeneity of $\mathcal{R}(\mathcal{K}, d)$.** An obvious weakness in Proposition 4.6 is that it is not explicit about when $\mathcal{R}(\mathcal{K}, d)$ is homogeneous. Note that we cannot provide a result analogous to Proposition 3.4 for homogeneity, because, for example, Dodgson's rule is known not to be homogeneous [11]. In [8, Section 4] some sufficient conditions for homogeneity of a rule were stated (without proof). We give some small generalizations here.

Definition 4.7. *Suppose that d is votewise and anonymous, and $\mathcal{K}$ is anonymous. Say that $(\mathcal{K}, d)$ satisfies the* **votewise minimizer property** *(VMP) if the following condition is satisfied.*

> *For each $r \in L_s(C)$ and each election $E = (C, V, \pi) \in \mathcal{E}$, there exists a minimizer $(C, V, \pi^*) \in \mathcal{K}_r$ of the distance from E to $\mathcal{K}_r$, so that for all $i, d(\pi_i, \pi_i^*)$ depends only on π_i and r.*

Remark 4.8. *The VMP implies that $d(E, \mathcal{K}_r)$ depends only on $\overline{E}$ and r. Furthermore $d(\pi_i, \pi_i^*)$ has the form $\delta(t, r)$ where $t, r \in L(C)$, and $d(E, \mathcal{K}_r)$ has the form $N(s)$ where s is the multiset of all values of $\delta(t, r)$, for $t \in L(C)$, counted with multiplicity.*

Proposition 4.9. *Let d be an anonymous votewise distance on $\mathcal{E}$.*

(i) Suppose that $\mathcal{K}$ is defined on Δ as follows: for each $r \in L(C)$, there is a nonempty subset S_r of $L(C)$ such that $\mathcal{K}_r = \{x | x_\rho = 0 \text{ for all } \rho \in S_r\}$. Then $(\mathcal{K}, d)$ satisfies the VMP.
(ii) $(\mathbf{W}, d)$ and $(\mathbf{S}, d)$ satisfy the VMP;
(iii) $(\mathbf{C}, d), (\mathbf{M}, d)$ and $(\mathbf{\Sigma}, d)$ do not necessarily satisfy the VMP.

Proof. For (i), the minimizer in question is obtained by, for each i, choosing the closest element of $L(C)$ under the underlying distance. Part (ii) follows immediately, because we can take $S_r = \{\rho \mid \rho \neq r\}$ for $\mathbf{S}$ and $S_r = \{\rho \mid \rho(1) \neq r(1)\}$ for $\mathbf{W}$. Finally, $(\mathbf{C}, d_H)$ and $(\mathbf{C}, d_K)$ do not satisfy the VMP. For example, consider the election $E = (C, V, \pi)$ where $C = \{a, b\}$, V has size 5, and $\pi = \{ab, ab, ba, ba, ba\}$. Then $d(E, \mathbf{C}_a) = 1$ for $d \in \{d_H, d_K\}$, and every minimizer differs from π only in that precisely one of the ba voters (say v_i) switches to ab. This also shows that $(\mathbf{M}, d)$ and $(\mathbf{\Sigma}, d)$ need not satisfy the VMP, because each coincides with $\mathbf{C}$ when $m = 2$. □

Proposition 4.10. *Suppose that $\mathcal{K}$ is homogeneous, d is votewise with respect to a homogeneous norm and $(\mathcal{K}, d)$ satisfies the VMP. Then $\mathcal{R}(\mathcal{K}, d)$ is homogeneous.*

Proof. Since $(\mathcal{K}, d)$ satisfies the VMP, we have that for all E and r, $d(E, K_r) = N(x)$ where $x = (d(\pi_1, r), ...d(\pi_n, r))$. Thus, for all k, $d(kE, K_r) = N(x^{(k)})$. By homogeneity of N, we have $d(E, K_r) = d(kE, K_r)$. So the winning rankings r in E and kE are the same and the rule is homogeneous. □

Corollary 4.11. *If d is votewise and based on the ℓ^p-norm for $1 \leq p \leq \infty$, then $\mathcal{R}(\mathbf{S}, d)$ and $\mathcal{R}(\mathbf{W}, d)$ are homogeneous. This generalizes [8, Thm 4.1]). Thus we recapture the well-known facts that positional scoring rules and Kemeny's rule are homogeneous, as is Copeland's rule.*

The case $p = \infty$ allows for formally stronger results, as in [8, Thm 4.6].

Corollary 4.11 shows, for example, that although Dodgson's rule can be rationalized with respect to $\mathbf{S}$ and some distance [10, Thm 3.3], no such distance can be homogeneous.

4.2. **The ℓ^p-votewise case - Wasserstein distance.** In the case of ℓ^p-votewise distances, we can relate $\overline{d}$ to a well-known concept from probability theory. Let S be a finite set and let $\Delta(S)$ denote the set of probability distributions on S. For a distance d over S, the function $d_W^p : \Delta(S) \times \Delta(S) \to \mathcal{R}$ is defined by

$$d_W^p(x, y)^p = \inf_A \sum_{r, r' \in S} A_{r,r'} d(r, r')^p,$$

where the infimum is taken over all couplings of x and y, defined as nonnegative square matrices of size $m!$ whose marginals are x and y respectively (i.e. $\forall r, \sum_{r'} A_{rr'} = x_r$ and $\forall r', \sum_r A_{rr'} = y_{r'}$). Basically, it represents the minimal cost to move from one configuration to another, where the underlying distance d defines the cost of each movement. Indeed, this construction leads to a new distance.

Proposition 4.12. *If d is a distance on S, then d_W^p is a distance on $\Delta(S)$. Moreover, if d is a pseudometric (respectively quasimetric or metric), then so is d_W^p.*

The function d_W^p is called the l^p**-transportation** distance, or p**-Wasserstein distance**. Now we are able to make the link with the votewise metrics.

Proposition 4.13. *Let d be an l^p-votewise distance. Then $\overline{d} = d_W^p$.*

Proof. Since d is homogeneous, $\overline{d}(x,y) = \min_{E'} d(E,E')$, where $\overline{E} = x$, $\overline{E'} = y$ and $|V(E)| = |V(E')| = n$, say. Let $E = (C,V,\pi)$ and $E' = (C,V,\pi')$. Then

$$\overline{d}(x,y)^p = \min_{\pi'} \frac{1}{n} \sum_i d(\pi_i, \pi'_i)^p = \min_a \frac{1}{n} \sum_{r,r' \in S} a_{r,r'} d(r,r')^p,$$

where the $a_{r,r'}$'s are integers such that for all $r \in S, \sum_r \frac{a_{r,r'}}{n} = x_r$ and for all $r' \in S, \sum_{r'} \frac{a_{r,r'}}{n} = y_{r'}$, which corresponds to the Wasserstein distance restricted to matrices A respecting the conditions and with coefficients of the form $\frac{k}{n}$ with $0 \leq k \leq n$. So clearly, $d_W^p(x,y) \leq \overline{d}(x,y)$.

Let assume that this inequality is strict: say the minimum in the definition of the Wasserstein distance is reached by a matrix A. Since $\max_{r,r'} d(r,r') < \infty$ and we can choose n as big as we want, A can be approximated arbitrarily close by a matrix A' of the previous form. We can get arbitrarily close to $d_W^p(x,y)$ in this way, which proves the equality. □

Example 4.14. *Let d_1 be the restriction to Δ of the ℓ^1 metric on $\mathbb{R}^M$. Then $\overline{d}_H = \frac{1}{2} d_1$ (also called the* **total variation distance***). This was observed (without the current notation) in [16, Lemma 3.2]: if E, E' are elections on (C,V) with $|V| = n$, then*

$$d_1(\overline{E}, \overline{E'}) \leq \frac{2}{n} d(E,E')$$

and given $\overline{E}, \overline{E} \in \Delta_{\mathbb{Q}}$, we can choose n and the preimages E, E' so that equality holds.

The 1-Wasserstein distance is also known as the **Earth Mover's** distance or **first Mallows metric** (used heavily in other areas of computer science, particularly image retrieval and pattern recognition).

5. Other properties of consensus sets

We are already restricting to consensuses in $\Delta_{\mathbb{Q}}$, that is, consensuses that satisfy anonymity and homogeneity. All the examples we have used so far have also satisfied *neutrality* and *convexity*.

We argue that convexity (defined in the usual way via restriction from $\mathbb{R}^M$) of each $\mathcal{K}_r$ is an essential condition. In the following example, it seems ridiculous that a should win at the extra point.

Example 5.1. *Consider the case $m = 3$, and the consensus formed by extending $\mathbf{W}$ so that a is the winner whenever $x_{bca} = x_{cba} = 1/2$ (and similarly for b, c). This consensus is anonymous and homogeneous, but $\mathcal{K}_a, \mathcal{K}_b, \mathcal{K}_c$ are not convex.*

In the simplex model, convexity is equivalent to the notion of **consistency**: if we split the voter set into two parts each of which elects r, the original voter set should elect r. It rules out the above example and we shall assume it from now on. Note that we only require each $\mathcal{K}_r$ to be convex, because convexity of the entire rule is a very restrictive assumption [23].

The group of permutations of C acts naturally on various sets. For example, it acts on $L(C)$ by $g(a_1, \ldots, a_m) = (g(a_1), \ldots, g(a_m))$ and on elections by $g(C,V,\pi) = (C,V,g(\pi))$ where $g(\pi) = (g(\pi_1), \ldots, g(\pi_n))$. An object is **neutral** if it is invariant under this action of G. Neutrality is a very natural condition for consensuses and for distances, and is satisfied by all our main examples.

Proposition 5.2. *Rule R is neutral if and only if R has the form $\mathcal{R}(\mathcal{K}, d)$ where $\mathcal{K}$ and d are neutral.*

Proof. Given a neutral R, the construction of Proposition 2.24 works. The converse is obvious from the definition of neutrality. □

Note that in general, by symmetry, there is no social choice function on (C, V) that is anonymous and neutral (for certain values of $|C|$ and $|V|$, such a function can exist). Thus in general we expect to have non-uniqueness of output from our social rule.

We suggest that when applying the DR framework, consensus classes should be required to satisfy (at least) anonymity, homogeneity, neutrality and convexity, while distances should be required to be metrics. Even these conditions do not rule out counterintuitive behaviour, as we see in Section 6.

6. Tied sets and decisiveness

All social rules used in practice encounter the problem of breaking ties. However, many commonly used social rules have the property that the tied subset is small (for example asymptotically negligible as $n \to \infty$, for fixed m). This is important, because if the tied region is small, then ties can be ignored for many purposes, whereas if the tied region is asymptotically large, our rule may suffer extreme lack of decisiveness. In this section we give "positive" (few ties) and negative (many ties) results.

The definitions in Section 2 fail to give a unique output precisely when there is some election E which is equidistant from at least two regions $\mathcal{K}_r$, each of which is closer to E than all others. The following definition also makes sense when we restrict to a given set or number of voters.

Definition 6.1. *The* **boundary** *of the rule of size s defined by $(\mathcal{K}, d)$ is the set of all elections at which the minimum in* (1) *is attained for at least two distinct $r \in L_s(C)$.*

Example 6.2. *Suppose that $m = 2$, with alternatives a and b, and that $\mathcal{K} = \mathbf{C}$ and d is an anonymous neutral standard distance. The concepts of majority winner and Condorcet winner coincide when $m = 2$. When n is odd, the boundary of the rule is empty, whereas when n is even, the profiles having an equal number of ab and ba voters are in the boundary. If d is homogeneous so that we define the rule on $\Delta_{\mathbb{Q}}$, then the point $(1/2, 1/2)$ is the unique element of the boundary.*

Example 6.3. *(large boundary) Consider the Copeland rule whose boundary contains all points with no unique Copeland winner. The sum of Copeland scores of all alternatives is $m(m-1)/2$. If m is odd, an m-way tie for first place is possible. The set of all such profiles corresponds to a subset of Δ having nonempty interior, so the boundary will certainly not be "small" in any reasonable sense. Since d_{RT} does not distinguish points, we might expect* a priori *that such a phenomenon is possible.*

Our intuition is that the boundary of a well-behaved homogeneous voting rule should be "small" in Δ, and be geometrically "nice". The relevant geometric theory is that of *Voronoi diagrams.* We now digress to review some known results, which will help develop a more refined intuition.

6.1. Geometric background.

All definitions below work for an arbitrary Minkowski space. For a fixed set of **sites** (subsets of the entire space), the open **Voronoi cell** of each site X is defined as the set of points closer to X than any other site. The boundaries of the cells are contained in the union of bisectors, where a **bisector** is the set of points equidistant from two sites and denoted $\beta(X, Y)$.

Interpreting the sites as the consensus sets $\mathcal{K}_r$, we see that the open Voronoi cell corresponding to $\mathcal{K}_r$ is precisely the set on which $(\mathcal{K}, d)$ is single-valued with value r. Also, the boundary as defined above is just the union of boundaries of all Voronoi cells, and this is contained in the union of all bisectors.

We first discuss bisectors, because if these are well-behaved, so will the boundary of the rule be (this is not a necessary condition, however).

Example 6.3 shows that the distance strongly influences the structure of bisectors. Thus we shall focus on the case of metrics from now on. We first restrict to the nicest situation, namely $\mathbb{R}^n$ under the Euclidean ℓ^2-norm,where our intuition is greatest.

We start with the simplest case where sites are single distinct points X and Y (when the sites are not single points, things can be more complicated even in ℓ^2). In Euclidean space, $\beta(X, Y)$ is a hyperplane normal to the line joining the points, and the Voronoi cells are therefore convex polyhedra that tile the entire space.

However, this situation is rather special. Consider a finite-dimensional normed space (**Minkowski space**). It is known, in fact, that $\beta(X, Y)$ is a hyperplane for all X and Y if and only if the space is ℓ^2 [4]. Thus we should expect to see bisectors that are not hyperplanes. Of course, such bisectors may still be "small" and "nice". Note that $\beta(X, Y)$ is homeomorphic to a hyperplane provided the norm is strictly convex and even sometimes when it is not [12] (unfortunately, this does not preclude nasty space-filling curves, which will fail to be "small" or "nice" under any reasonable definitions).

We conclude that small nice bisectors should not be expected in general. In the next section we give some special cases.

6.2. **Small bisectors and hyperplane rules.** All our results in this section show that the bisectors in question are contained in a finite union of hyperplanes. Rules which have a well-defined winner on each component of the complement in Δ of a finite set of hyperplanes have been studied recently. Note that such rules are automatically anonymous and homogeneous. Mossel, Procaccia and Racz [16] call them **hyperplane rules** and show their equivalence with the **generalized scoring rules** of Xia and Conitzer [22]. These rules can be defined axiomatically using **finite local consistency** [21].

Most rules that have ever been studied by social choice theorists are hyperplane rules. A notable exception is Copeland's rule. In order to interpret Copeland's rule as a hyperplane rule, [16] requires that the winner be (arbitrarily) specified on the tied region. This seems to us to be stretching the definition too far.

In [16] it was shown that for a hyperplane rule, the minimum size of a coalition that can change the winner, under the uniform distribution on $\mathcal{P}$, follows a universal law. Many results in [16] will extend to the case of rules with "small" boundary.

We now give a sufficient condition for a DR rule to be a hyperplane rule.

Proposition 6.4. *Let d be ℓ^p-votewise for some $1 \le p < \infty$ and suppose that $(\mathcal{K}, d)$ satisfies the VMP. Then on Δ, $\beta(\mathcal{K}_r, \mathcal{K}_{r'})$ is defined by*

$$\sum_{t \in L(C)} x_t \left[\delta(t, r)^p - \delta(t, r')^p\right] = 0.$$

Proof. We know that the distance between $E = (C, V, \pi)$ and the minimizer $m(E, r) = (C, V, \pi^*)$ equals $N(s)$ where s is the multiset with entries $\delta(t, r)$ occurring according to their multiplicities nx_t, for all $t \in L(C)$. The specific form of N then shows that $d(E, m(E, r))^p = n\left(\sum_t x_t \delta(t, r)^p\right)$. Applying the same argument for r' and rearranging shows that the bisector is defined by the displayed equation. □

Corollary 6.5. *Suppose that d is ℓ^p-votewise with $1 \le p < \infty$, and $\mathcal{K}$ is defined as in Proposition 4.9. Then $\mathcal{R}(\mathcal{K}, d)$ is a hyperplane rule.*

Proof. By Proposition 4.9(i), VMP is satisfied. It suffices to show that the linear function in Proposition 6.4 is not identically zero. The distance from x to $\mathcal{K}_r$ is attained at a point $m(x, r)$ where $m(x, r)_t = x_t$ for all $t \notin S$, and $d(x, \mathcal{K}_r) = \sum_{t \in S} x_t \delta(t, r)^p$. If $r \ne r'$ then by definition $S_r \ne S_{r'}$. Thus taking $t \notin S \cap S'$, without loss of generality $\delta(t, r) = 0$ and $\delta(t, r') \ne 0$. □

Corollary 6.6. *Every rule of the form $\mathcal{R}(\mathbf{S}, d)$ or $\mathcal{R}(\mathbf{W}, d)$, where d is ℓ^p-votewise and $1 \le p < \infty$, is a hyperplane rule.*

Remark 6.7. *This result does not extend to general distances. For example, Copeland's rule as we have defined it is not a hyperplane rule, having a large tied set in general, yet it can be defined as $\mathcal{R}(\mathbf{W}, d_{RT})$.*

Note that when $p = \infty$, we do not obtain a hyperplane rule. In fact, every point $x \in \Delta_{\mathbb{Q}}$ for which every coordinate is nonzero is equidistant from all $\mathbf{S}_r$, so the rule is almost maximally indecisive.

6.3. **Large bisectors — analysis of ℓ^1-votewise metrics.** Despite the positive results above, we already know enough to expect negative results. Large bisectors occur often when the unit ball in a Minkowski space is not strictly convex. We now show in Proposition 6.9 that this geometric property holds for all ℓ^1-votewise metrics.

We have already seen that d_H leads to the restriction of the usual ℓ^1-metric to $\Delta_{\mathbb{Q}}$. That metric is induced by a norm and hence has some useful properties, in particular translation-invariance. We first generalize this result.

Let $c = (1, 1, \dots, 1)/M \in \mathbb{R}^M$ be the center of the simplex. Now, we translate the center c to the origin, and we denote by Δ' the image of the simplex under this translation. We denote by $\mathcal{H}$ the hyperplane containing Δ'. Our study of the geometry under the Wasserstein distance will be facilitated by the following observation.

Proposition 6.8. *Let d be a distance. Then d_W^p is translation-invariant.*

Proof. For any vector v, let $v^+ = (\max(v_1, 0), \dots, \max(v_n, 0))$. We show that for all integer p and for all points x and y, $d_W^p(x, y) = d_W^p((x-y)^+, (y-x)^+)$, which implies that d_W is translation-invariant.

Clearly, $d_W^p(x, y) \le d_W^p((x-y)^+, (y-x)^+)$. To show that $d_W(x, y) = d_W((x-y)^+, (y-x)^+)$, we show that there exists A such that $d_w(x, y)^p = \sum_{r \ne r'} A_{r,r'} d(r, r')^p$ and $\forall r, A_{r,r} = \min(x_r, y_r)$.

Assume that it is not the case, and choose A reaching the minimum with a maximal trace (we always have $A_{rr} \le \min(x_r, y_r)$ because of the marginals condition): there exists r such that $A_{rr} < \min(x_r, y_r)$. By the marginals condition, it implies that there also exists r' and r'' such that $A_{r,r'}$ and $A_{r'',r}$ are strictly positive. Then, define the nonnegative square matrices A' of size $m!$ whose marginals are x and y respectively, and such that $A'_{r,r} = A_{r,r} + \min(A_{r,r'}, A_{r'',r})$, $A'_{r'',r'} = A_{r'',r'} + \min(A_{r,r'}, A_{r'',r})$, and for all other couples c different from (r', r) and (r'', r), $A'_c = A_c$. It verifies $\sum_{r \ne r'} A'_{r,r'} d(r, r')^p \le \sum_{r \ne r'} A_{r,r'} d(r, r')^p$ because $d(r'', r') \le d(r'', r) + d(r, r')$. Since A' has a strictly bigger trace than A, this contradicts our original assumption. □

Restricting to ℓ^1, we can obtain a normed space.

Proposition 6.9. *Let d be an ℓ^1-votewise distance. Then*

(i) d induces a norm N on $\mathcal{H}$. Explicitly, for each $x \in \Delta'$, we have $N(x) = d_W^1(x + c, c)$;
(ii) the unit ball of the norm N is not strictly convex.

Proof. By Proposition 6.8 it suffices to show that the transportation metric is homogeneous (in the usual sense of the norm property) over Δ', and as we just said, N is its homogeneous extension to $\mathcal{H}$. We want to show that for any $x \in \Delta'$ and λ such that $\lambda x \in \Delta'$, $d_W^1(\lambda x + c, c) = |\lambda| d_W^1(x + c, c)$. We fix now λ and x. Let A be a matrix achieving the minimum equal to $d_W^1(x + c, c)$ and satisfying the corresponding conditions. Let A' be the matrix with marginals $\lambda x + c$ and c such that, for $r \ne r'$, $A'_{rr'} = \lambda A_{rr'}$ if $\lambda > 0$, and $|\lambda| ({}^{\mathrm{t}}A)_{rr'}$ otherwise: because of the conditions on the marginals, we get $\forall r, A'_{rr} = \min(\lambda x_r + c_r, c_r)$. The matrix A' is nonnegative if and only if $\lambda x \in \Delta'$: it is nonnegative if and only for all r, $\min(\lambda x_r + c_r, c_r) \ge 0$ which occurs if and only if $\lambda x + c$ is in the simplex. In addition, $\sum_{r,r' \in S} A'_{r,r'} d_S(r, r')^p = |\lambda| \sum_{r,r' \in S} A_{r,r'} d_S(r, r')^p$, so we obtain $d_W^1(\lambda x + c, c) \ge |\lambda| d_W^1(x + c, c)$.

In the other hand, $x = \frac{1}{\lambda}\lambda x$, and so, by the same argument, we have that $d^1_W(\lambda x + c, c) \leq |\lambda| d^1_W(x + c, c)$: we have the equality, and N is a norm, yielding (i).

For (ii), fix a ranking $r \in L(C)$ and consider the subset S_r of all points x where only the component corresponding to r is negative. In S_r, we have

$$N(x) = \sum_{r'} x_{r'} d^1_W(r', r).$$

Thus the intersection of S_r with the unit sphere is contained in a hyperplane and hence not strictly convex. □

Remark 6.10. *This result is not true for the other Wasserstein metrics, because they do not satisfy the homogeneity property of norms. For example, if $x \in \Delta'$ is such that $x_r \geq 0$ and $\forall r \neq r', x_{r'} \leq 0$, then $d^p_W(x+c, c)^p = \sum_{r'} |x_{r'}| d(r, r')^p$, and $d^p_W(\lambda x + c, c) = |\lambda|^{\frac{1}{p}} \sum_{r'} |x_{r'}| d(r, r')^p = |\lambda|^{\frac{1}{p}} d^p_W(x + c, c)$.*

We can now show that large bisectors occur in the DR framework. We say that a subset of a Minkowski space is *large* if it contains an open ball, and *small* otherwise.

Example 6.11. *(large bisector) Let $m = 3$ and consider the line segment L_1, L_2 which join the centre of Δ to the points $x_{abc} = 1$ and $x_{bca} = 1$ respectively. Let $\mathcal{K}_a = L_1, \mathcal{K}_b = L_2$ and $\mathcal{K}_c$ be the single point $x_{cab} = 1$. Every point in the cone spanned by the normals to L_1, L_2, and making an acute angle with at least one of these normals, lies on the bisector of L_1, L_2 with respect to the usual ℓ^2-metric. Every point in this set that is closer to the centre of Δ than to $\mathcal{K}_c$ is in the boundary of the rule, which is therefore large. Note that $\mathcal{K}$ satisfies strong unanimity, anonymity, homogeneity, but not neutrality. Furthermore, every $\mathcal{K}_r$ is a convex polyhedral subset of the simplex.*

Presumably the same behaviour occurs with many other metrics. Assuming neutrality makes it harder to construct such examples.

We now show that even for consensuses that consist of isolated points, large bisectors can occur.

Proposition 6.12. *Consider a norm N induced over $\mathcal{H}$ by an ℓ^1-votewise metric. Let r, r_1, r_2 be rankings. We denote by d_1 and d_2 the distances $d(r, r_1)$ and $d(r, r_2)$. Let $\epsilon > 0$. We define x and y as the two points of $\mathcal{H}$ such that $x_r = -x_{r_1} = \frac{\epsilon}{d_1}$, $y_r = -y_{r_2} = \frac{\epsilon}{d_2}$ and all other components are equal to zero. Then, any point $z \in \mathcal{H}$ such that $z_r \geq 1 - \frac{\epsilon}{\min(d_1, d_2)}$ is equidistant from x and y according to N.*

Proof. Let z be such a point. Then, $x - z$ and $y - z$ have only one negative component: the one corresponding to the ranking r. So $N(x - z) = \sum_{r'} (x_{r'} - z_{r'}) d(r, r')$ and $N(y - z) = \sum_{r'} (y_{r'} - z_{r'}) d(r, r')$. Since the only components where x and y differ are r_1 and r_2, they are equidistant of z if and only if $(x_{r_1} - z_{r_1}) d_1 + (x_{r_2} - z_{r_2}) d_2 = (y_{r_1} - z_{r_1}) d_1 + (y_{r_2} - z_{r_2}) d_2$, which is equivalent to $x_{r_1} d_1 + x_{r_2} d_2 = y_{r_1} d_1 + y_{r_2} d_2$, which is in turn equivalent to $x_{r_1} d_1 = y_{r_2} d_2$, which is always true by definition. □

Corollary 6.13. *Let d be an l^1-votewise metric. Then there is a consensus $\mathcal{K}$ consisting of isolated points, so that the boundary of $\mathcal{R}(\mathcal{K}, d)$ is large.*

Proof. Write x_ϵ and y_ϵ for points x and y of the form defined in Proposition 6.12. That proposition implies that, by setting $\mathcal{K}_a = \{x_\epsilon\}$ and $\mathcal{K}_b = \{y_\epsilon\}$ and choosing a sufficiently small ϵ, then $\beta(\mathcal{K}_a, \mathcal{K}_b)$ will be large. Also, for any other candidate c, if we set $\mathcal{K}_c = \{x_{\epsilon'}\}$ with $\epsilon < \epsilon'$, then for any z such that $z_r \geq 1 - \frac{\epsilon}{\min(d_1, d_2)}$, $N(z, \mathcal{K}_a) = N(z, \mathcal{K}_b) < N(z, \mathcal{K}_c)$. □

Remark 6.14. *Note that the consensus in the proof of Corollary 6.13 is somewhat unnatural. For example, it is not neutral.*

The next question is how often this kind of situation happens. For simplicity we focus on the case of ℓ^1. We can give an exact characterization of when two points have a large bisector. This is directly connected with the well-known integer partition problem.

Proposition 6.15. *Let $x, y \in \mathbb{R}^M$. We denote by S the set of values $(x_i - y_i)$. Then x and y have a large bisector under ℓ^1 if and only if there exists a subset $S' \subset S$ such that $\sum_{e \in S'} e = \sum_{e \notin S'} e$.*

Proof. By definition $\beta(x, y) = \{z | \sum_i |x_i - z_i| = \sum_i |y_i - z_i|\}$. We divide $\mathbb{R}^M$ into 4^M subspaces corresponding to the possible signs of the the values $(x_i - z_i)$ and $(y_i - z_i)$. Let V be one of these subspaces: in V, the equality $\sum_i |x_i - z_i| = \sum_i |y_i - z_i|$ is equivalent to $\sum_i \epsilon_i(x_i - z_i) = \sum_i \epsilon'_i(y_i - z_i)$, where $\forall i, \epsilon_i, \epsilon'_i = \pm 1$. There are two cases. First, if $\exists i, \epsilon_i \neq \epsilon'_i$ then the equation can be put into the form $x_i = k_j x_j + C$, where $k_j \in \{-2, 0, 2\}$ and some $k_j \neq 0$ and so this part of $\beta(x, y)$ lies in a hyperplane and is therefore small. The other case is when $\forall i, \epsilon_i = \epsilon'_i$. Then $V \cap \beta(x, y)$ is large if and only if $\sum_i \epsilon_i(x_i - y_i) = 0$, which is equivalent to the fact that there exists $S' \subset S, \sum_{e \in S'} e = \sum_{e \notin S'} e$. □

We can now characterize further, if it exists, the shape of the large bisector of two points: any ball included in the bisector is contained in a subset defined by a set of equations $z_i \leq \min(x_i, y_i)$ or $z_i \geq \max(x_i, y_i)$ for all i. It implies, for example, that if the points are corners of the simplex, the large bisector in $\mathbb{R}^M$ does not intersect Δ. Thus, for example, large bisectors cannot occur with $\mathbf{S}$ (which we already know from Corollary 6.5).

Definition 6.16. *Define the decision problem LARGE-BISECTOR as follows. Input is a pair (x, y) of integer points of $\mathcal{R}^M$ and we must decide whether $\beta(x, y)$ contains an open ball in ℓ^1. Note that M is part of the input.*

Corollary 6.17. *LARGE-BISECTOR is NP-complete, and so the analogous decision problem for rational points of Δ is NP-hard.*

Proof. We use Proposition 6.15. Given an instance (x, y, M) of LARGE-BISECTOR, let $z_i = x_i - y_i$, giving an instance of the known NP-complete problem PARTITION. Given an instance $(z_1, \ldots, z_M)$ of PARTITION, let $x_i = z_i, y_i = 0$. This gives an instance of LARGE-BISECTOR. In each case the original instance is a yes instance of the associated decision problem if and only if the derived instance is. □

We suspect the analogue of LARGE-BISECTOR to be NP-hard for every l^1-votewise metric.

The situation is quite subtle, because large bisectors do not occur when the consensus sets are hyperplanes instead of points.

Proposition 6.18. *The bisector of two distinct hyperplanes under any norm on $\mathbb{R}^n$ is contained in a union of at most two hyperplanes.*

Proof. The distance from a point x to a hyperplane H defined by $a^T x = b$ is equal to $d(x, H) = \frac{|a^T x - b|}{||a||^*}$ where * denotes the dual norm (see for example [14]). Now, let H' be another hyperplane defined by the equation $a'x = b'$. We assume that $||a||^* = ||a'||^*$ (without loss of generality since multiplying by a scalar still defines the same hyperplane). The bisector of H and H' can be defined as the set of points x verifying $|a^T x - b| = |a'^T x - b'|$. So, we have two cases, depending on the sign of these absolute values: either $\sum_i (a_i - a'_i)x_i = b - b'$ or $\sum_i (a_i + a'_i)x_i = b + b'$. Since $H \neq H'$, each of these is the equation of a hyperplane. □

Even when large bisectors do occur, the boundary of the rule is often small. We do not know of any "natural" rule based on a votewise metric for which the boundary is large.

7. Future work

Systematic exploration of rules $(\mathcal{K}, d)$, where $\mathcal{K}$ satisfies the properties suggested at the end of Section 5 and d is an ℓ^p-votewise metric, may prove fruitful in finding new rules with desirable properties. Even the rules defined by $(\mathbf{S}, d)$ and $(\mathbf{W}, d)$, where d is ℓ^p-votewise, have not been thoroughly explored, to our knowledge. They are hyperplane rules, which is a major point in their favour as far as decisiveness goes. After projection to the *permutahedron* (which is just changing the underlying distance in our framework), the cases $p = 2$ (*mean proximity rules*) [13] and $p = 1$ (*mediancenter rules*) [3] have received attention.

Some important questions, some of which partially motivated the present work, remain unanswered.

- The question of whether Dodgson's rule is a hyperplane rule has not been settled here or in [16]. Is there a natural DR rule, defined with respect to a votewise metric and a homogeneous consensus, that is not a hyperplane rule?
- What are good sufficient conditions on $\mathcal{K}$ and d in Δ to guarantee that the boundary of $\mathcal{R}(\mathcal{K}, d)$ is small? For example, is some combination of the following assumptions sufficient: $\mathcal{K}$ and d are neutral, d is a votewise metric, the sets $\mathcal{K}_r$ are separated, and $(\mathcal{K}, d)$ satisfies the VMP?
- We have only a few sufficient conditions for homogeneity of a DR rule. In [11] a way around the nonhomogeneity of Dodgson's rule was found, by using a homogenization process. How is the homogenization of a DR rule defined in general?

There are some more issues that may be of interest.

- If we use other distances on Δ, not derived from distances on profiles, do we find anything interesting? Note that **statistical distances** such as Kullback-Liebler distance are heavily used in many areas.
- In [19] a related notion of convexity was investigated. In that case, the authors worked directly on profiles and only with the Kemeny distance. Their notion can be generalized immediately to any shortest path distance. This notion of convexity differs from ours, but the precise relationship between them is not clear to us and should be further studied.
- Although the Wasserstein p-distance is not induced by a norm in general, perhaps the topology is similar enough that we can derive results to those in the normed case.
- The hypothesis that distances be standard can probably be weakened in several results.

References

[1] D.E. Campbell and S.I. Nitzan. "Social compromise and social metrics". English. *Social Choice and Welfare* 3. (1986), 1–16.

[2] Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. "Modal ranking: A uniquely robust voting rule". In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*. 2014, pp. 616–622.

[3] Davide P Cervone et al. "Voting with rubber bands, weights, and strings". *Mathematical Social Sciences* 64. (2012), 11–27.

[4] Mahlon M. Day. "Some characterizations of inner-product spaces". *Trans. Amer. Math. Soc.* 62 (1947), 320–337.

[5] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistic, 1988.

[6] E Elkind, P Faliszewski, and A Slinko. "Rationalizations of Condorcet-consistent rules via distances of hamming type". *Social Choice and Welfare* 39. (2012), 891–905.

[7] Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. "Good Rationalizations of Voting Rules". In: *AAAI*. Ed. by Maria Fox and David Poole. AAAI Press, 2010.

[8] Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. "Homogeneity and monotonicity of distance-rationalizable voting rules". In: *AAMAS*. Ed. by Liz Sonenberg et al. IFAAMAS, 2011, pp. 821–828.

[9] Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. "On distance rationalizability of some voting rules". In: *TARK*. Ed. by Aviad Heifetz. 2009, pp. 108–117.

[10] Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. "On the role of distances in defining voting rules". In: *AAMAS*. Ed. by Wiebe van der Hoek et al. IFAAMAS, 2010, pp. 375–382.

[11] P.C. Fishburn. “Condorcet Social Choice Functions”. *SIAM Journal on Applied Mathematics* 33 (1977), 469–489.

[12] A.G. Horvath. “On bisectors in Minkowski normed spaces”. *Acta Mathematica Hungarica* 89 (2000), 233–246.

[13] Sébastien Lahaie and Nisarg Shah. “Neutrality and geometry of mean voting”. In: *Proceedings of the fifteenth ACM conference on Economics and computation*. ACM. 2014, pp. 333–350.

[14] O. L. Mangasarian. “Arbitrary-Norm Separating Plane”. *Operations Research Letters* 24 (1997), 15–23.

[15] Tommi Meskanen and Hannu Nurmi. “Closeness Counts in Social Choice”. English. In: *Power, Freedom, and Voting*. Ed. by Matthew Braham and Frank Steffen. Springer Berlin Heidelberg, 2008, 289–306.

[16] Elchanan Mossel, Ariel D. Procaccia, and Miklós Z. Rácz. “A smooth transition from powerlessness to absolute power”. *Journal of Artificial Intelligence Research* (2013), 923–951.

[17] James Munkres. “Algorithms for the assignment and transportation problems”. *Journal of the Society for Industrial and Applied Mathematics* 5. (1957), 32–38.

[18] Shmuel Nitzan. “Some measures of closeness to unanimity and their implications”. English. *Theory and Decision* 13. (1981), 129–138.

[19] Svetlana Obraztsova et al. “On swap-distance geometry of voting rules”. In: *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS ’13, Saint Paul, MN, USA, May 6-10, 2013*. Ed. by Maria L. Gini et al. IFAAMAS, 2013, pp. 383–390.

[20] Donald G. Saari. *Geometry of voting*. Vol. 3. Studies in Economic Theory. Springer-Verlag, Berlin, 1994, pp. xvi+372.

[21] Lirong Xia and Vincent Conitzer. “Finite Local Consistency Characterizes Generalized Scoring Rules”. In: *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*. Ed. by Craig Boutilier. 2009, pp. 336–341.

[22] Lirong Xia and Vincent Conitzer. “Generalized scoring rules and the frequency of coalitional manipulability”. In: *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*. Ed. by Lance Fortnow, John Riedl, and Tuomas Sandholm. ACM, 2008, pp. 109–118.

[23] H. Young. “Social Choice Scoring Functions”. *SIAM Journal on Applied Mathematics* 28. (1975), 824–838. eprint: `http://epubs.siam.org/doi/pdf/10.1137/0128067`.

ENS Lyon
E-mail address: `benjamin.hadjibeyli@ens-lyon.fr`

University of Auckland
E-mail address: `mcw@cs.auckland.ac.nz`