Handbook of Computational Social Choice

Edited by Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang,
and Ariel Procaccia

Illustrations

Tables

8
Rationalizations of Voting Rules

8.1 Introduction

From antiquity to these days, voting has been an important tool for making collective decisions that accommodate the preferences of all participants. Historically, a remarkably diverse set of voting rules have been used (see, e.g., Brams and Fishburn, 2002), with several new voting rules proposed in the last three decades (Tideman, 1987; Schulze, 2003; Balinski and Laraki, 2010). Thus, when decision-makers need to select a voting rule, they have plenty of choice: should they aggregate their opinions using something as basic as Plurality voting or something as sophisticated as Ranked Pairs? Or should they perhaps design a new voting rule to capture the specific features of their setting?

Perhaps the best known way to answer this question is to use the axiomatic approach, i.e., identify desirable properties of a voting rule and then choose (or construct) a rule that has all of these properties. This line of work was initiated by Arrow (1951) and led to a great number of impossibility theorems, as it turned out that some desirable properties of voting systems are incompatible. By relaxing these properties, researchers obtained axiomatic characterizations of a number of classical voting rules, such as Majority (May, 1952), Borda (Young, 1975) and Kemeny (Young and Levenglick, 1978); see the survey by Chebotarev and Shamis (1998) as well as Chapter 2 (Zwicker, 2015).

However, early applications of voting suggest a different perspective on this question. It is fair to say that in the Middle Ages voting was most often used by religious organizations (Uckelman and Uckelman, 2010). The predominant view in ecclesiastical elections was that God's cause needed the most consecrated talent that could be found for leadership in the church. Moreover, it was believed that God knew who the best candidate was, so the purpose of elections was to reveal God's will. It is therefore not surprising that when the Marquis de Condorcet (1785) undertook the first attempt at systematization of voting rules, he was influenced by the philosophy of church elections. His view was that the aim of voting is to determine the "best"

[a] Department of Computer Science, University of Oxford, United Kingdom
[b] Department of Mathematics, University of Auckland, New Zealand

decision for the society when voters are prone to making mistakes. This approach assumes that there is an objectively correct choice, but voters have different opinions due to errors of judgment; absent these errors, they would all agree on the most suitable alternative. Thus, one should aim to design a voting rule that maximizes the probability of identifying the best choice. Depending on the model of "noise" or "mistakes" in voters' judgment, we get different voting rules. In statistics, this approach is known as *maximum likelihood estimation (MLE)*: it tries to estimate the state of the world (which is hidden) that is most likely to produce the observed noisy data.

A somewhat different, but related approach, which takes its roots in ideas of Charles Dodgson (1876), can be called consensus-based. The society agrees on a notion of a consensus (for example, we could say that there is a consensus if all voters agree which alternative is the best, or if there exists a Condorcet winner), and the result of each election is viewed as an imperfect approximation to a consensus. Specifically, if a preference profile R is a consensus, then we pick the consensus winner, and otherwise we output the winners of consensus profiles R' that are as close to R as possible. Alternatively, we may say that the society looks for a minimal change to the given preference profile that turns it into a profile with an indisputable winner. At the heart of this approach is the agreement as to (1) which preference profiles should be viewed as consensual and (2) what is the appropriate notion of closeness among preference profiles. It turns out that many common voting rules can be explained and classified by different choices of these parameters.

In this chapter we will survey the MLE framework and the consensus-based framework, starting with the latter. We demonstrate that both frameworks can be used to rationalize many common voting rules, with the consensus-based framework being somewhat more versatile. We also establish some connections between the two frameworks. We remark that these two frameworks are not the only alternatives to the axiomatic analysis. For instance, Camps et al. (2014) put forward an approach that is based on propositional logic. Further, in economic literature the term "rationalization" usually refers to explaining the behavior of an agent or a group of agents via an acyclic (or transitive) preference relation, and there is a large body of literature that investigates which voting rules are rationalizable in this sense (see Bossert and Suzumura, 2010, for a survey). In this chapter, we focus on the MLE framework and the consensus-based framework because these two methods for rationalizing voting rules are interesting from a computational perspective: as we will see, explaining a voting rule via a consensus and a "good" measure of closeness implies upper bounds on its algorithmic complexity, whereas MLE-based voting rules are desirable for many applications, such as crowdsourcing, and therefore implementing them efficiently is of paramount importance.

In what follows, we assume that the set of alternatives is A and $|A| = m$; we use the terms *alternatives* and *candidates* interchangeably. Also, unless specified otherwise, voters' preferences and ballots are assumed to be linear orders over A.

8.2 Consensus-Based Rules

The goal of the consensus-based approach is to reach a compromise among all voters, i.e., to arrive at a situation where there is agreement in society as to which outcome is the best. This may require persuading some voters to modify their opinions in minor ways, and, as a result, to make small changes to their ballots. Obviously it is desirable to minimize the number and magnitude of these changes. Thus, the best alternatives are the ones for which the agreement can be reached at the smallest cost (measured by the total amount of changes). In other words, given an arbitrary preference profile, we proceed by identifying the consensual profiles that are most similar to it and outputting their winners. The result then depends on how we define consensual profiles and how we measure the magnitude of change in votes. The latter question is usually addressed by using a distance over the space of profiles; this is why voting rules that can be obtained in this manner are called *distance rationalizable*. Often, this distance is obtained by computing the number of "unit changes" needed to transform one profile into the other, where the notion of "unit change" may vary from one voting rule to another.

This method of constructing voting rules can be traced back to Dodgson (1876), who was the first to define a voting rule in this manner (for a specific notion of consensus and a specific distance between profiles, see Section 8.2.1). More recently, it was formalized and studied by Nitzan (1981), Lerer and Nitzan (1985), Campbell and Nitzan (1986), and Baigent (1987), and subsequently by Meskanen and Nurmi (2008) and Elkind et al. (2010a,b, 2011, 2012); we also point the reader to the survey of Eckert and Klamler (2011). It turns out that many classic voting rules can be obtained in this manner; Meskanen and Nurmi (2008) put together an extensive catalogue of distance rationalizations of common voting rules, with additional examples provided by Elkind et al. (2010b, 2012). Furthermore, many properties of voting rules can be derived from their distance rationalizations: a voting rule can be shown to have "nice" properties if it can be rationalized via a "nice" consensus class and a "nice" distance. This makes the distance rationalizability approach eminently suitable for constructing new voting rules: it allows us to combine known distances and consensus classes, and derive conclusions about the resulting rules based on the properties of their components.

We start by presenting a few examples that illustrate the concepts of consensus and distance to consensus, followed by a formal definition and a discussion of properties of distance rationalizable voting rules.

8.2.1 Examples

The examples in this section are taken from the work of Meskanen and Nurmi (2008) and Elkind et al. (2012); see these papers for additional references. We provide brief descriptions of the voting rules we consider; for formal definitions the reader is referred to Chapter 2 (Zwicker, 2015).

Dodgson. Perhaps the most canonical example of the consensus-based approach is the Dodgson rule. Recall that winner determination under this rule proceeds as follows. If the given preference profile has a Condorcet winner, i.e., a candidate that beats every other candidate in a pairwise election, then this candidate is declared the unique Dodgson winner. Otherwise, for every candidate c we compute her *Dodgson score*, i.e., the number of swaps of adjacent candidates in voters' ballots that need to be performed in order to make c a Condorcet winner. We then output all candidates with the smallest Dodgson score. This definition follows the principles of the distance rationalizability framework: the underlying notion of agreement is the existence of a Condorcet winner, and the unit changes are swaps of adjacent candidates. This notion of unit change corresponds to a distance on rankings known as the *swap distance*, which is the number of swaps of adjacent candidates needed to transform one ranking into the other. We refer the reader to Chapter 5 (Caragiannis et al., 2015) for a complexity-theoretic analysis of the Dodgson rule.

Kemeny. The Kemeny rule is also defined in terms of the swap distance. While it is more common to view this rule as a *social preference function*, i.e., a mapping that, given a preference profile, outputs a set of rankings, in this section we will be interested in the interpretation of this rule as a social choice function. Under the Kemeny rule, we identify all rankings that minimize the total swap distance to the voters' ballots. The associated social preference function then outputs all such rankings, whereas the Kemeny social choice function (which we will refer to as the Kemeny rule) outputs all candidates that are ranked first in at least one of these rankings. This rule can be viewed as another example of the distance rationalizability approach: the consensual profiles are ones where all votes are identical, and the unit changes are the same as for the Dodgson rule, i.e., swaps of adjacent candidates.

Plurality. Under Plurality rule, each candidate gets one point from each voter who ranks her first; the winners are the candidates with the largest number of points. Since Plurality considers voters' top candidates only, it is natural to use a notion of consensus that also has this property: we say that there is an agreement in the society if all voters rank the same candidate first. Now, consider an n-voter preference profile. If some candidate a receives $n_a \leq n$ Plurality votes, there are $n - n_a$ voters who *do not* rank her first. Thus, if we want to turn this profile into a consensus where everyone ranks a first, and we are allowed to change the ballots in any way we like (at a unit cost per ballot), we have to modify $n - n_a$ ballots. In other words, if our notion of a unit change is an arbitrary modification of an entire ballot, then the number of unit changes required to make a candidate a consensus winner is inversely related to her Plurality score. In particular, the candidates for whom the number of required unit changes is minimal are the Plurality winners. Alternatively, we can define a unit change as a swap of two (not necessarily adjacent) candidates; the argument above still applies, thereby showing that this construction also leads to the Plurality rule.

Borda. Recall that the Borda score of a candidate a in an n-voter, m-candidate profile is given by $(m - r_1) + \cdots + (m - r_n) = nm - \sum_i r_i$, where r_i, $i = 1, \ldots, n$, is the rank of a in the i-th ballot. To distance rationalize this rule, we use the same notion of consensus as for the Plurality rule (i.e., all voters agree on who is the best candidate) and the same notion of unit change as for the Dodgson rule and the Kemeny rule, namely, a swap of adjacent candidates. Indeed, to ensure that a is ranked first by voter i, we need to perform $r_i - 1$ swaps of adjacent candidates. Consequently, making a the unanimous winner requires $\sum_i r_i - n$ swaps. That is, the number of swaps required to make a candidate a consensus winner is inversely related to her Borda score. This construction, which dates back to Farkas and Nitzan (1979), can be extended to scoring rules other than Borda, by assigning appropriate weights to the swaps (Lerer and Nitzan, 1985).

Copeland. The Copeland score of a candidate a can be defined as the number of pairwise elections that a wins (a may also get additional points for the pairwise elections that end in a tie; in what follows we focus on elections with an odd number of voters to avoid dealing with ties). The Copeland winners are the candidates with the highest Copeland score. For this rule, an appropriate notion of consensus is the existence of a Condorcet winner. As for the notion of unit change, it is convenient to formulate it in terms of the pairwise majority graph. Recall that the pairwise majority graph $\mathcal{G}(R)$ of a profile R over a candidate set A is the directed graph whose vertex set is A and there is a directed edge from candidate a to candidate b if a strict majority of voters in R prefer a to b. Consider two n-voter profiles R^1 and R^2 over a candidate set A; assume that n is odd. A natural notion of a unit change in this setting is an edge reversal, i.e., a pair $(a, b) \in A \times A$ such that in $\mathcal{G}(R^1)$ there is an edge from a to b, whereas in $\mathcal{G}(R^2)$ there is an edge from b to a. The distance between R^1 and R^2 is then defined as the number of edge reversals. To see that this distance combined with the Condorcet consensus rationalizes the Copeland rule, note that if a candidate's Copeland score is s, she can be made the Condorcet winner by reversing $m - 1 - s$ edges, so the number of edge reversals and the candidate's Copeland score are inversely related.

Maximin. The Maximin score of a candidate a in an n-voter profile R over a candidate set A is the number of votes that a gets in her most difficult pairwise election (i.e., $\min_{b \in A} n_{ab}$, where n_{ab} is the number of voters in R who prefer a to b); the winners are the candidates with the highest score. Suppose that R has no Condorcet winner, and consider a candidate $a \in A$. Let b be a's most difficult opponent, i.e., a's Maximin score is $s_a = n_{ab}$; note that $s_a \leq \frac{n}{2} < \frac{n+1}{2}$, since a is not a Condorcet winner. Then if we add $n+1-2s_a$ ballots where a is ranked first, a will be the Condorcet winner in the resulting profile (which has $2n+1-2s_a$ voters, with $n+1-s_a$ of these voters ranking a above c for every $c \in A$). On the other hand, if we add $k < n+1-2s_a$ ballots, we obtain a profile where at least $n - s_a$ voters out of $n+k$ prefer b to a; as $2(n - s_a) \geq n + k$, this means that at least half of the voters in this profile prefer b to a, so a is not a Condorcet winner. Thus, a candidate's Maximin score is inversely related to the number of ballots that need to be added in order to obtain a profile where this candidate is the Condorcet winner.

This argument explains the Maximin rule in the language of agreement and changes. However, this explanation does not quite fit our framework, since it uses a notion of unit change (adding a single ballot) that does not directly correspond to a distance. The problem here is that a distance is supposed to be symmetric (see Section 8.2.2), whereas adding ballots is an inherently asymmetric operation: if we can turn R into R' by adding s ballots, we cannot turn R' into R by adding s ballots. It turns out, however, that the Maximin rule can be rationalized via the distance that measures the number of ballots that need to be *added or deleted* to turn one profile into another (see Elkind et al., 2012, for details). Intuitively, this is because for the purpose of reaching a Condorcet consensus adding a ballot is always at least as useful as deleting a ballot.

We remark that there is another voting rule that is defined in terms of deleting ballots so as to obtain a Condorcet consensus, namely the Young rule, which is discussed in Chapter 5 (Caragiannis et al., 2015). While the Young rule, too, can be distance-rationalized, the construction is quite a bit more complicated than for Maximin (Elkind et al., 2012).

These examples raise a number of questions. First, is it the case that all voting rules can be explained within the consensus-based framework? Second, what are the appropriate notions of consensus and distance to consensus? Third, can we derive any conclusions about a voting rule based on the notion of consensus and distance that explain it? To answer these questions, we need to define our framework formally.

8.2.2 Formal Model

The consensus-based framework that has been introduced informally so far has two essential components: the definition of what it means to have an agreement in the society and the notion of distance between preference profiles. We will now discuss both of these components in detail. Our presentation mostly follows Elkind et al. (2010b).

Consensus Classes

Informally, we say that a preference profile R is a consensus if it has an undisputed winner reflecting a certain concept of agreement in the society. Formally, a *consensus class for a set of candidates* A is a pair $\mathcal{K} = (\mathcal{X}, w)$ where $\mathcal{X}$ is a nonempty set of profiles over A and $w\colon \mathcal{X} \to A$ is a mapping that assigns a unique candidate to each profile in $\mathcal{X}$; this candidate is called the *consensus choice (winner)*.[1] We require $\mathcal{K}$ to be anonymous and neutral, in the following sense: For every profile $R \in \mathcal{X}$ a profile R' obtained from R by permuting voters satisfies $R' \in \mathcal{X}$ and $w(R') = w(R)$, and the profile R'' obtained from R by renaming candidates according to a permutation $\pi : A \to A$ satisfies $R'' \in \mathcal{X}$ and $w(R'') = \pi(w(R))$ (i.e., the winner under R'' is obtained by renaming the winner under R according to π).

The following classes of preference profiles have been historically viewed as situations of consensus:

Strong unanimity. This class, denoted $\mathcal{S}$, consists of profiles where all voters report the same preference order. The consensus choice is the candidate ranked first by all voters. The reader may note that we have used this notion of consensus in Section 8.2.1 to rationalize the Kemeny rule. Interestingly, it can also be used to provide an alternative rationalization of the Plurality rule (Elkind et al., 2010a).

Unanimity. This class, denoted $\mathcal{U}$, consists of profiles where all voters rank some candidate c first (but may disagree on the ranking of the remaining candidates). The consensus choice is this candidate c. This consensus class appears in our rationalizations of Plurality and Borda. It is also used to rationalize other scoring rules (Lerer and Nitzan, 1985; Elkind et al., 2009).

Majority. This class, denoted $\mathcal{M}$, consists of profiles where more than half of the voters rank some candidate c first. The consensus choice is this candidate c. This notion of consensus can be used to rationalize Plurality and a simplified version of the Bucklin rule (Elkind et al., 2010b).

Condorcet. This class, denoted $\mathcal{C}$, consists of profiles with a Condorcet winner. The consensus choice is the Condorcet winner. This notion of consensus

[1] One can also consider situations in which the voters reach a consensus that several candidates are equally well qualified to be elected; this may happen, for example, under Approval voting when all voters approve the same set of candidates. However, in what follows we limit ourselves to consensus classes with unique winners.

appears in our rationalizations of the Dodgson rule, the Copeland rule, and Maximin.

Transitivity. This class, denoted $\mathcal{T}$, consists of profiles whose majority relation is transitive, i.e., for every triple of candidates $a, b, c \in A$ it holds that if a majority of voters prefer a to b and a majority of voters prefer b to c, then a majority of voters prefer a to c. Such profiles always have a Condorcet winner, so we define the consensus choice to be the Condorcet winner. This consensus class can be used to rationalize the Slater rule (Meskanen and Nurmi, 2008).

It is easy to see that we have the following containment relations among the consensus classes: $\mathcal{S} \subset \mathcal{U} \subset \mathcal{M} \subset \mathcal{C}$ and $\mathcal{S} \subset \mathcal{T} \subset \mathcal{C}$. However, $\mathcal{U}$ and $\mathcal{T}$ are incomparable, i.e., $\mathcal{U} \not\subseteq \mathcal{T}$ and $\mathcal{T} \not\subseteq \mathcal{U}$. Similarly, we have $\mathcal{M} \not\subseteq \mathcal{T}$ and $\mathcal{T} \not\subseteq \mathcal{M}$.

Remark 8.1 A consensus class $(\mathcal{X}, w)$ can be viewed as a voting rule with domain $\mathcal{X}$ that always outputs a unique candidate. Conversely, every anonymous and neutral voting rule f such that $|f(R)| = 1$ for at least one profile R defines a consensus class: if f is defined on the set of all profiles over a candidate set A, we can define a consensus class $\mathcal{K}_f = (\mathcal{X}_f, w_f)$ by setting $\mathcal{X}_f = \{R \mid |f(R)| = 1\}$ and for each $R \in \mathcal{X}_f$ defining $w_f(R)$ to be the unique candidate in $f(R)$. That is, this consensus class consists of all profiles on which f makes a definitive choice. The condition that $|f(R)| = 1$ for some profile R is necessary to ensure that $\mathcal{X}_f \neq \emptyset$.

There are other consensus classes one could consider: for example, one could study a 2/3-variant of the majority consensus $\mathcal{M}$, where more than 2/3 of the voters rank the same candidate first (this choice of threshold stems from the observation that in many countries changes to the constitution require the support of two thirds of the eligible voters). However, these five classes appear to be representative enough to rationalize many interesting voting rules.

Distances

To capture the idea of measuring the magnitude of changes in a preference profile, we use distances on profiles. Recall that a *distance* on a set X is a mapping $d\colon X \times X \to \mathbb{R} \cup \{+\infty\}$ such that for every $x, y, z \in X$ the following four conditions are satisfied:

(a) $d(x, y) \geq 0$ (nonnegativity);
(b) $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles);
(c) $d(x, y) = d(y, x)$ (symmetry);
(d) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

A mapping that satisfies (a), (c), and (d), but not (b), is called a *pseudodistance.*

For distance rationalizability constructions, we need distances that are defined on pairs of profiles. Usually, it is enough to only consider pairs of profiles with the same set of candidates (this will be the case for all distances considered in this chapter), and in many cases it suffices to only consider pairs of profiles with the same number of voters. In particular, to construct a distance on the space of all n-voter profiles over a fixed set of candidates A, we can take a suitable distance d on the space $\mathcal{L}(A)$ of all linear orders over A and extend it to a distance $\widehat{d}$ over the space of all n-voter preference profiles $\mathcal{L}^n(A)$ by setting

$$\widehat{d}((u_1, \ldots, u_n), (v_1, \ldots, v_n)) = d(u_1, v_1) + \ldots + d(u_n, v_n). \tag{8.1}$$

It can be shown that $\widehat{d}$ satisfies all distance axioms whenever d does. This method of building distances over profiles from distances over votes will play an important role in our analysis (see Section 8.2.4).

We will now present several examples of distances on the space of preference profiles. Some of these distances should look familiar to the reader, as they were used to rationalize voting rules in Section 8.2.1.

Discrete distance. The *discrete distance* is defined on pairs of profiles with the same set of candidates A and the same number of voters n using formula (8.1); the underlying distance on $\mathcal{L}(A)$ is given by

$$d_{\text{discr}}(u, v) = \begin{cases} 0 & \text{if } u = v, \\ 1 & \text{if } u \neq v. \end{cases}$$

This distance was used in our rationalization of the Plurality rule.

Swap distance. The *swap distance*, which is also known as the Kendall tau distance, the Kemeny distance, the Dodgson distance, and the bubble-sort distance (Kendall and Gibbons, 1990), is also defined using formula (8.1). The underlying distance on $\mathcal{L}(A)$ is the swap distance between individual votes: $d_{\text{swap}}(u, v)$ is the number of pairs $(c, c') \in A \times A$ such that u ranks c above c', but v ranks c' above c.

(Weighted) footrule distance. This distance is also known as Spearman distance, or Spearman footrule (Kendall and Gibbons, 1990). Let $\text{pos}(u, c)$ denote the position of candidate c in vote u (the top candidate in u has position 1, and the bottom candidate in u has position m). Then the footrule distance on $\mathcal{L}(A)$ is given by

$$d_{\text{fr}}(u, v) = \sum_{c \in A} |\text{pos}(u, c) - \text{pos}(v, c)|.$$

That is, we measure the displacement of each candidate as we move from u to v, and then we take the sum over all candidates. This distance is extended to preference profiles using formula (8.1). The reader can verify that we can use the footrule distance $\widehat{d}_{\text{fr}}$ instead of the swap distance in our rationalization of the Borda rule.

Further, let $\alpha = (\alpha_1, \dots, \alpha_m)$ be a vector of m nonnegative rationals (weights). We define a (pseudo)distance $d_{\text{fr-}\alpha}(u, v)$ on $\mathcal{L}(A)$ by setting

$$d_{\text{fr-}\alpha}(u, v) = \sum_{c \in A} |\alpha_{\text{pos}(u,c)} - \alpha_{\text{pos}(v,c)}|. \tag{8.2}$$

When all weights are distinct, $d_{\text{fr-}\alpha}$ is a distance. However, when some of the weights coincide, $d_{\text{fr-}\alpha}$ is a pseudodistance, but not a distance. The reader can verify that for $\alpha = (m-1, \dots, 1, 0)$ the distance $d_{\text{fr-}\alpha}$ coincides with d_{fr}. It can be shown that by using $\widehat{d}_{\text{fr-}\alpha}$ we can rationalize the scoring rule with the score vector α (Elkind et al., 2009), i.e., the rule that, given a profile $R = (v_1, \dots, v_n)$, outputs the set $\text{argmax}_{a \in A}(\alpha_{\text{pos}(v_1,a)} + \dots + \alpha_{\text{pos}(v_n,a)})$.

ℓ_∞-Sertel distance. This distance, denoted by $\widehat{d_{\text{sert}}}^{\infty}$, is also obtained by extending a distance on rankings to n-voter profiles; however, in contrast with all distances considered so far, it is not defined via formula (8.1). Let $u(i)$ denote the candidate ranked in position i in vote u. We define the distance $d_{\text{sert}} : \mathcal{L}(A) \times \mathcal{L}(A) \to \mathbb{R}$ by setting

$$d_{\text{sert}}(u, v) = \max\{i \mid u(i) \neq v(i)\},$$

with the convention that $d_{\text{sert}}(u, v) = 0$ if $u = v$. The ℓ_∞-Sertel distance on n-voter preference profiles is then defined by setting

$$\widehat{d_{\text{sert}}}^{\infty}((u_1, \dots, u_n), (v_1, \dots, v_n)) = \max_{i=1,\dots,n} d_{\text{sert}}(u_i, v_i).$$

The reason for having the symbol ℓ_∞ in the name of this distance and the notation $\widehat{d_{\text{sert}}}^{\infty}$ will become clear in Section 8.2.4. This distance, together with the majority consensus, can be used to provide a rationalization of a simplified version of the Bucklin rule (Elkind et al., 2010b).

Edge reversal (pseudo)distance. This distance is defined over the set of all profiles with an odd number of voters. Given two profiles R^1, R^2 over A, we set

$$d_{\text{rev}}(R^1, R^2) = |\{(a, b) \in A \times A \mid a >_{R^1} b, b >_{R^2} a\}|,$$

where we write $a >_R b$ to denote that a majority of voters in the profile R prefer a to b. This distance counts the number of edges in the pairwise majority graph of R^1 that need to be reversed to obtain the pairwise majority graph of R^2. The edge reversal distance was used in our rationalization of the Copeland rule; it can also be used to rationalize the Slater rule (Meskanen and Nurmi, 2008).

Note that, technically speaking, d_{rev} is a pseudodistance rather than a distance: we have $d_{\text{rev}}(R^1, R^2) = 0$ whenever R^1 and R^2 have the same pairwise majority graph. It is perhaps more natural to think of the domain of d_{rev} as the space of all tournaments over A, in which case d_{rev} satisfies all distance axioms.

Vote insertion (pseudo)distance. This distance is also defined over the set of all profiles with a given candidate set. Consider two profiles R^1 and R^2 over a candidate set A whose multisets of votes are given by V^1 and V^2 respectively. The vote insertion distance d_{ins} between R^1 and R^2 is the size of the symmetric difference between V^1 and V^2. This distance computes the cost of transforming R^1 into R^2 (or vice versa) if we are allowed to add or delete votes at a unit cost. Elkind et al. (2012) show that by combining this distance with the Condorcet consensus we obtain the Maximin rule. Again, d_{ins} is a pseudodistance rather than a distance: $d_{\text{ins}}(R^1, R^2) = 0$ if R^1 and R^2 have the same multiset of votes. It can be viewed as a distance on the space of *voting situations*, i.e., multisets of votes over A.

We are now ready to put together the two components of our framework.

Definition 8.2 Let d be a (pseudo)distance on the space of preference profiles over a candidate set A, and let $\mathcal{K} = (\mathcal{X}, w)$ be a consensus class for A. We define the $(\mathcal{K}, d)$*-score* of a candidate a in a profile R to be the distance (according to d) between R and a closest profile $R' \in \mathcal{X}$ such that a is the consensus winner of R'. The set of $(\mathcal{K}, d)$*-winners* in a profile R consists of all candidates in A whose $(\mathcal{K}, d)$-score is the smallest.

Definition 8.3 A voting rule f is *distance rationalizable* via a consensus class $\mathcal{K}$ and a distance d over profiles (or, $(\mathcal{K}, d)$*-rationalizable*) if for every profile R a candidate is an f-winner in R if and only if she is a $(\mathcal{K}, d)$-winner in R.

We can now formalize our analysis of the six examples in Section 8.2.1: our arguments show that the Dodgson rule is $(\mathcal{C}, \widehat{d}_{\text{swap}})$-rationalizable, the Kemeny rule is $(\mathcal{S}, \widehat{d}_{\text{swap}})$-rationalizable, Plurality is $(\mathcal{U}, \widehat{d}_{\text{discr}})$-rationalizable, the Borda rule is $(\mathcal{U}, \widehat{d}_{\text{swap}})$-rationalizable, the Copeland rule is $(\mathcal{C}, d_{\text{rev}})$-rationalizable, and Maximin is $(\mathcal{C}, d_{\text{ins}})$-rationalizable. Observe that three of these well-known voting rules can be rationalized using the same distance (but different consensus classes). Further examples can be found in the work of Nitzan (2010): Chapter 6 of his book provides a summary of rules that are rationalizable with respect to the unanimity consensus. Meskanen and Nurmi (2008) describe distance rationalizations for several other voting rules; while some of these rationalizations are very appealing, others appear less intuitive. Motivated by this observation, we will now try to formalize what it means to have a "good" distance rationalization.

8.2.3 Universal Distance Rationalizability

It turns out that the unrestricted distance rationalizability framework defined in Section 8.2.2 is too powerful: Lerer and Nitzan (1985) show that if we do not impose any restrictions on the distance used, then essentially *any* voting rule is rationalizable with respect to all the standard consensus classes. This result was subsequently rediscovered by Elkind et al. (2010b), and our presentation follows their work.

To formally state this universal distance rationalizability result, we need a notion of compatibility between a voting rule and a consensus class.

Definition 8.4 A voting rule f is said to be *compatible* with a consensus class $\mathcal{K} = (\mathcal{X}, w)$, or $\mathcal{K}$-*compatible*, if $f(R) = \{w(R)\}$ for every profile R in $\mathcal{X}$.[2]

We will now show that every voting rule is distance rationalizable with respect to every consensus class that it is compatible with.

Theorem 8.5 *Let A be a set of candidates, let f be a voting rule over A, and let $\mathcal{K} = (\mathcal{X}, w)$ be a consensus class for A. Then f is $(\mathcal{K}, d)$-rationalizable for some distance d if and only if it is $\mathcal{K}$-compatible.*

Proof Let f be a voting rule that is $(\mathcal{K}, d)$-rationalizable for some consensus class $\mathcal{K} = (\mathcal{X}, w)$ and distance d. Let R be some profile in $\mathcal{X}$. There is only one profile at distance 0 from R—namely, R itself. Hence, the unique $(\mathcal{K}, d)$-winner in R is $w(R)$. Thus, f is $\mathcal{K}$-compatible.

Conversely, suppose that f is $\mathcal{K}$-compatible. We will now define a distance d over the set of all profiles over the candidate set A as follows. We set $d(R, R') = 0$ if $R = R'$. We set $d(R, R') = 1$ if (a) $R \in \mathcal{X}$ and $w(R) \in f(R')$ or (b) $R' \in \mathcal{X}$ and $w(R') \in f(R)$. In all other cases, we set $d(R, R') = 2$. It is easy to check that d satisfies all distance axioms. It remains to argue that f is $(\mathcal{K}, d)$-rationalizable.

Consider a profile $R \in \mathcal{X}$. Since f is $\mathcal{K}$-compatible, we have $f(R) = \{w(R)\}$. Furthermore, we have $d(R, R) = 0$ and there is no profile R', $R' \neq R$, such that $d(R, R') = 0$. Thus, the unique $(\mathcal{K}, d)$-winner in R is $w(R)$, too.

On the other hand, consider a profile $R \notin \mathcal{X}$. Note that $d(R, R') \geq 1$ for every profile $R' \in \mathcal{X}$. Since $\mathcal{K}$ is neutral and $\mathcal{X} \neq \emptyset$, for each $a \in f(R)$ there exists a consensus profile R^a in which a is the consensus winner. By construction, we have $d(R, R^a) = 1$. Further, we have $d(R, R') = 2$ for every profile $R' \in \mathcal{X}$ such that $w(R') \notin f(R)$. Thus, the set $f(R)$ is exactly the set of $(\mathcal{K}, d)$-winners in R, and the proof is complete. □

Theorem 8.5 implies that being compatible with any of our five standard consensus classes suffices for distance rationalizability. Now, almost all common voting rules are compatible with the strong unanimity consensus $\mathcal{S}$, and hence distance rationalizable. This argument does not apply to voting rules that do not have unique winners on strongly unanimous profiles, such as Veto and k-Approval for $k > 1$

[2] One might think that the term "$\mathcal{K}$-consistent" would be more appropriate than "$\mathcal{K}$-compatible." Indeed, a voting rule that elects the Condorcet winner whenever one exists is usually referred to as Condorcet-consistent. We chose to use the term "$\mathcal{K}$-compatible" to avoid confusion with the normative axiom of consistency.

(recall that k-Approval is the scoring rule with the score vector $(\underbrace{1,\dots,1}_{k},\underbrace{0,\dots,0}_{m-k})$, and Veto is simply the $(m-1)$-Approval rule). However, both Veto and k-Approval can be shown to be distance rationalizable by a slightly different argument.

Corollary 8.6 *For every anonymous neutral voting rule f over a set of candidates A such that $|f(R)| = 1$ for some profile R there exist a consensus class $\mathcal{K} = (\mathcal{X}, w)$ and a distance d such that f is $(\mathcal{K}, d)$-rationalizable.*

Proof We can use the consensus class $\mathcal{K}_f = (\mathcal{X}_f, w_f)$ defined in Remark 8.1: by definition, f is $\mathcal{K}_f$-compatible, so Theorem 8.5 implies that f is $(\mathcal{K}_f, d)$-rationalizable for some distance d. □

Clearly, both Veto and k-Approval satisfy the conditions of Corollary 8.6, so they are distance rationalizable as well.

Yet, intuitively, the distance used in the proof of Theorem 8.5 is utterly unnatural. For instance, we have seen that the Dodgson rule and the Kemeny rule can be rationalized via the swap distance, which is polynomial-time computable. In contrast, Elkind et al. (2010b) show that applying Theorem 8.5 to either of these rules results in a rationalization via a distance that is not polynomial-time computable (assuming P $\neq$ NP)—this follows from the fact that winner determination for these rules is computationally hard, as discussed in Chapter 5 (Caragiannis et al., 2015).

Thus, knowing that a rule is distance rationalizable—even with respect to a standard notion of consensus—by itself provides no further insight into the properties of this rule; for a rationalization to be informative, the distance used must be natural. Consequently, we will now shift our focus from distance rationalizability *per se* to quality of rationalizations, and seek an appropriate subclass of distances that would be expressive enough to capture many interesting rules while allowing us to draw nontrivial conclusions about rules that they rationalize.

8.2.4 Votewise Distances

In this section, we focus on distances that are obtained by first defining a distance on preference orders and then extending it to profiles. The reader may observe that the distances $\widehat{d}_{\text{discr}}$, $\widehat{d}_{\text{swap}}$, $\widehat{d}_{\text{fr}}$, and $\widehat{d_{\text{sert}}}^{\infty}$ defined in Section 8.2.2 are constructed in this way. This class of distances was identified by Elkind et al. (2010b), and our presentation in this section is based on their work.

Definition 8.7 A *norm* on $\mathbb{R}^n$ is a mapping $N\colon \mathbb{R}^n \to \mathbb{R}$ that has the following properties:

(a) positive scalability: $N(\alpha u) = |\alpha| N(u)$ for all $u \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$;
(b) positive semidefiniteness: $N(u) \geq 0$ for all $u \in \mathbb{R}^n$, and $N(u) = 0$ if and only if $u = (0, 0, \dots, 0)$;

(c) triangle inequality: $N(u+v) \leq N(u) + N(v)$ for all $u, v \in \mathbb{R}^n$.

A well-known class of norms on $\mathbb{R}^n$ is that of *p-norms* ℓ_p, $p \in \mathbb{Z}^+ \cup \{\infty\}$, given by

$$\ell_p(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \text{ for } p \in \mathbb{Z}^+, \quad \ell_\infty(x_1, \ldots, x_n) = \max\{|x_1|, \ldots, |x_n|\}.$$

In particular, $\ell_1(x_1, \ldots, x_n) = |x_1| + \cdots + |x_n|$.

Definition 8.8 Let A be a fixed set of candidates, fix $n > 0$, let d be a distance on $\mathcal{L}(A)$, and let N be a norm on $\mathbb{R}^n$. We say that a distance D on the space of n-voter profiles over the candidate set A is *N-votewise* if for every pair of profiles R and R' over A with $R = (u_1, \ldots, u_n)$ and $R' = (v_1, \ldots, v_n)$ we have

$$D(R, R') = N(d(u_1, v_1), \ldots, d(u_n, v_n)). \tag{8.3}$$

It is easy to check that for every distance d on $\mathcal{L}(A)$ and every norm N on $\mathbb{R}^n$ the function defined by (8.3) is a (pseudo)distance. We will denote this (pseudo)distance by $\widehat{d^N}$. If $N = \ell_p$ for some $p \in \mathbb{Z}^+ \cup \{\infty\}$, we will write $\widehat{d^p}$ instead of $\widehat{d^{\ell_p}}$. Further, since many distance rationalizations use ℓ_1 as the underlying norm, we will write $\widehat{d}$ instead of $\widehat{d^1}$ (note that this notation is consistent with the one used earlier in this chapter for $\widehat{d}_{\text{swap}}$, $\widehat{d}_{\text{discr}}$, $\widehat{d}_{\text{fr}}$, and $\widehat{d_{\text{sert}}^{\infty}}$).

Given a norm N, we say that a rule is *N-votewise* if it can be distance rationalized via an N-votewise distance; we say that a rule is *votewise* if it is N-votewise for some norm N.

Votewise distances are expressive enough to rationalize many classic voting rules. For instance, the rationalizations of the Dodgson rule, the Kemeny rule, Plurality, and the Borda rule described in Section 8.2.1 demonstrate that all these rules are ℓ_1-votewise, and the results of Lerer and Nitzan (1985) and Elkind et al. (2009, 2010b) imply that the class of votewise rules includes essentially all scoring rules,[3] a simplified version of the Bucklin rule, and several other less common voting rules.

We will now demonstrate that votewise rules have a number of desirable properties, both from a normative and from a computational perspective.

Normative Properties of Votewise Rules

An important feature of the votewise distance rationalizability framework is that one can derive properties of votewise rules from the properties of their components, i.e., the underlying distance on votes, the norm, and the consensus class. Elkind et al. (2010b, 2011) consider such classic normative properties of voting rules as anonymity, neutrality, continuity, consistency, homogeneity and monotonicity, and,

[3] The exceptions are rules like Veto, which are not compatible with any standard consensus class; however, even such rules are votewise rationalizable via a pseudodistance.

for each of them, derive sufficient conditions on the components of a votewise rationalization for the resulting rule to have the respective property. We present a sample of these results below.

Anonymity. Recall that a voting rule is said to be *anonymous* if its result does not change when the ballots are permuted. It turns out that anonymity of a votewise rule is inherited from the corresponding norm. Specifically, a norm N on $\mathbb{R}^n$ is said to be *symmetric* if it satisfies $N(x_1, \ldots, x_n) = N(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for every permutation σ of $\{1, \ldots, n\}$; note that all p-norms are symmetric. Elkind et al. (2010b) show the following easy result.

Proposition 8.9 *Suppose that a voting rule f is $(\mathcal{K}, \widehat{d^N})$-rationalizable for some pseudodistance d over $\mathcal{L}(A)$, a consensus class $\mathcal{K}$, and a symmetric norm N. Then f is anonymous.*

Neutrality. A voting rule is said to be *neutral* if its result does not depend on the candidates' names. Neutrality of a votewise rule is a property of the underlying distance on votes. Namely, a distance d on $\mathcal{L}(A)$ is said to be *neutral* if for every permutation $\pi : A \to A$ and every pair of votes $u, v \in \mathcal{L}(A)$ it holds that $d(u, v) = d(u', v')$ where u' and v' are obtained from, respectively, u and v by renaming the candidates according to π. The following proposition is due to Elkind et al. (2010b).

Proposition 8.10 *Suppose that a voting rule f is $(\mathcal{K}, \widehat{d^N})$-rationalizable for some norm N, a consensus class $\mathcal{K}$, and a neutral pseudodistance d over $\mathcal{L}(A)$. Then f is neutral.*

Consistency. A voting rule f is said to be *consistent* if for every pair of profiles R^1, R^2 such that $f(R^1) \cap f(R^2) \neq \emptyset$, the preference profile $R^1 + R^2$ obtained by concatenating R^1 and R^2 satisfies $f(R^1 + R^2) = f(R^1) \cap f(R^2)$. Consistency is a very demanding property: while all common voting rules are anonymous and neutral, the class of voting rules that are anonymous, neutral and consistent consists of compositions of scoring rules (Young, 1975). Nevertheless, Elkind et al. (2010b) obtain a sufficient condition for a distance rationalizable voting rule to be consistent.

Proposition 8.11 *Suppose that a voting rule f is $(\mathcal{U}, \widehat{d^p})$-rationalizable for some $p \in \mathbb{Z}^+$ and some pseudodistance d over $\mathcal{L}(A)$. Then f is consistent.*

Homogeneity. A voting rule f is said to be *homogeneous* if for every profile R and every positive integer k it holds that $f(R) = f(kR)$, where kR is the preference profile obtained by concatenating k copies of R. This notion can be seen as a relaxation of the notion of consistency. Elkind et al. (2011) present several sufficient conditions for homogeneity of a votewise rule. For instance, they show that many of the voting rules that can be rationalized via the ℓ_∞ norm are homogeneous.

Proposition 8.12 *Suppose that a voting rule f is $(\mathcal{K}, \widehat{d^\infty})$-rationalizable for some* *pseudodistance d over $\mathcal{L}(A)$ and a consensus class $\mathcal{K} \in \{\mathcal{S}, \mathcal{U}, \mathcal{M}\}$. Then f is homogeneous.*

Monotonicity. A voting rule f is said to be *monotone* if moving a winning candidate upwards in some voters' preference orders (without changing the relative order of other candidates) does not make him a loser. To identify sufficient conditions for monotonicity of a votewise rule, Elkind et al. (2011) introduce several notions of monotonicity for distances over votes. In particular, they define *relatively monotone* distances. These are the distances over $\mathcal{L}(A)$ such that for every candidate $a \in A$ the following condition holds. Suppose that we have:

(i) two votes $y, y' \in \mathcal{L}(A)$ such that y and y' rank all candidates in $A \setminus \{a\}$ in the same order, but y' ranks a higher than y does, and
(ii) two votes $x, z \in \mathcal{L}(A)$ such that x ranks a first and z does not.

Then

$$d(x, y) - d(x, y') \geq d(z, y) - d(z, y'). \tag{8.4}$$

Elkind et al. (2011) show that the relative monotonicity condition is satisfied by the swap distance. Moreover, they prove the following result.

Proposition 8.13 *Suppose that a voting rule f is $(\mathcal{K}, \widehat{d})$-rationalizable for some relatively monotone distance d over $\mathcal{L}(A)$ and a consensus class $\mathcal{K} \in \{\mathcal{S}, \mathcal{U}\}$. Then f is monotone.*

Algorithmic Properties of Votewise Rules

Votewise rules are also appealing from a complexity-theoretic perspective: it turns out that we can show tractability results for them under a mild condition on the underlying distance. For the definitions of the complexity classes mentioned in this section, we refer the reader to the book of Hemaspaandra and Ogihara (2002).

Definition 8.14 We say that a distance D on the space of profiles over a candidate set A is *normal* if:

(a) D is polynomial-time computable;
(b) D takes values in the set $\mathbb{Z}^+ \cup \{+\infty\}$;
(c) if R^1 and R^2 have a different number of votes, then $D(R^1, R^2) = +\infty$.

Given a voting rule f, we consider the problem of determining whether a given candidate is one of the winners in a given profile under f; we refer to this problem as f-WINNER. Elkind et al. (2010b) show the following set of results for this problem. Suppose that a voting rule f is $(\mathcal{K}, D)$-rationalizable for some normal distance D and a consensus class $\mathcal{K} \in \{\mathcal{S}, \mathcal{U}, \mathcal{M}, \mathcal{C}\}$. Then:

(i) f-WINNER is in P^{NP};

(ii) if there exists a polynomial p such that for every pair of n-voter m-candidate profiles R^1, R^2 it holds that $D(R_1, R_2) \leq p(m+n)$, then f-Winner is in Θ_2^p;
(iii) if there exists a distance on votes d such that $D = \widehat{d}$, then f-Winner is in FPT with respect to the number of candidates;
(iv) if there exists a distance on votes d such that $D = \widehat{d}$ or $D = \widehat{d^\infty}$, $\mathcal{K} \in \{\mathcal{U}, \mathcal{M}\}$, and D is neutral, then f-Winner is in P/poly.

The first two results extend to the transitivity consensus $\mathcal{T}$ (which was not considered by Elkind et al. (2010b)); note also that for these results the distance D is not required to be votewise. However, it is not clear if the FPT algorithm in (iii) can be extended to $\mathcal{T}$ as well.

We emphasize that it is not the case that for every votewise rule the winner determination problem is in P (unless P = NP). In fact f-Winner may be intractable even if f is ℓ_1-votewise rationalizable with respect to a standard consensus class via an easy-to-compute distance on votes: examples are provided by the Dodgson rule and the Kemeny rule, which are known to be computationally hard (Hemaspaandra et al., 1997, 2005).

Votewise Distances: Discussion

We have seen that many common voting rules admit votewise distance rationalizations, and that distance rationalizable voting rules have several desirable properties. On the other hand, the "trivial" distance rationalization presented in Theorem 8.5 is clearly not votewise. Furthermore, some voting rules (most notably, STV) can be shown not to admit a votewise distance rationalization with respect to the standard consensus classes (Elkind et al., 2010a); we remark that the known distance rationalization for STV (Meskanen and Nurmi, 2008) is rather complex. Thus, the concept of a votewise distance appears to be useful for distinguishing between "good" and "bad" rationalizations.

Note, however, that the rationalizations of the Copeland rule and Maximin given in Section 8.2.1 are not votewise, despite being quite simple and intuitive. In fact, it is not known whether these rules are votewise distance rationalizable. It remains a challenge to come up with a definition of a "good" distance rationalization that covers all intuitively appealing rationalizations, but excludes the rationalization described in Theorem 8.5.

8.3 Rules as Maximum Likelihood Estimators

We will now turn our attention to voting rules that can be represented as maximum likelihood estimators. We start by revisiting the probabilistic model put forward by Condorcet (1785), and its interpretation by Young (1988).

Briefly, the basic assumption of Condorcet's model is that there always exists a correct ranking of the alternatives, which, however, cannot be observed directly.

Voters derive their preferences over the alternatives from this ranking: when comparing two alternatives, each voter is more likely to make the correct judgment than the incorrect one. Moreover, voters make their decisions independently from each other, and *a priori* each ranking is equally likely to be correct.

Formalizing Condorcet's ideas turned out to be a challenging task; below, we discuss some of the reasons for this. However, from a historical perspective, his ideas are very important, as they represent one of the earliest applications of what is now known as the *maximum likelihood estimation* approach. Under this approach, one computes the likelihood of the given preference profile for each possible "state of the world", i.e., the true ranking of the alternatives. The best ranking(s) of the alternatives are then the one(s) that have the highest likelihood of producing the given profile. If we assume a uniform prior over the space of all possible rankings, this procedure can be interpreted as estimating the most likely state of the world given the preference data (the equivalence of the two interpretations follows immediately from the Bayes rule).

Condorcet's approach can be extended in two different directions: First, we can consider different *noise models*, i.e., ways in which voters' preferences may arise from the true state of the world. Second, instead of associating a state of the world with a ranking of the alternatives, we can associate it with the identity of the best alternative (or, more generally, a set of pairwise comparisons between the alternatives); this approach is particularly attractive if the goal is to determine a single election winner rather than a full ranking of the alternatives (and in particular if there is indeed a unique "correct solution" to the decision problem at hand). Below we survey recent research that explores these directions.

8.3.1 Two Alternatives: Condorcet Jury Theorem

When there are only two alternatives to choose from, it is natural to use *majority voting*, i.e., select an alternative that is supported by at least half of the voters (breaking ties arbitrarily). It turns out that this is also the right strategy in Condorcet's model; in fact, as the number of voters grows, the probability that majority voting identifies the better alternative approaches 1. This result is known as the *Condorcet Jury Theorem*, and dates back to the original paper of Condorcet (1785).

Theorem 8.15 *Suppose that $|A| = 2$, and a priori each of the alternatives in A is equally likely to be the better choice. Suppose also that there are n voters, and each voter correctly identifies the better alternative with probability p, $1/2 < p \leq 1$; further, each voter makes her judgment independently from the other voters. Then the probability that the group makes the correct decision using the simple majority rule approaches 1 as $n \to +\infty$.*

Theorem 8.15 follows immediately from the Chernoff bound (see, e.g., Alon and Spencer, 2008); Condorcet's proof was based on a direct combinatorial argument.

Theorem 8.15 can be extended in a variety of ways. For instance, it can be generalized to the case where voters are *a priori* not identical, i.e., voter i's probability to make the correct choice is p_i and not all p_is are equal: Nitzan and Paroush (1982) and Shapley and Grofman (1984) show that in this case it is optimal to use weighted voting, assigning a weight of $\log \frac{p_i}{1-p_i}$ to voter i. However, in practice the probabilities p_i are often not known; to mitigate this, Baharad et al. (2011, 2012) propose a procedure for estimating them. Other extensions deal with settings where voters are not independent (see, e.g., Shapley and Grofman, 1984; Berg, 1993a,b; Ladha, 1992, 1993, 1995; Dietrich and List, 2004) or strategic (Austen-Banks and Smith, 1994; McLennan, 1998; Peleg and Zamir, 2012), or *a priori* the alternatives are not symmetric and the voters' probabilities of making the correct choice depend on the state of nature (Ben-Yashar and Nitzan, 1997).

When $|A| > 2$, the analysis becomes more complicated. In particular, it depends on whether the goal is to identify the most likely ranking of alternatives or the alternative that is most likely to be ranked first. We will now consider both of these options, starting with the former.

8.3.2 Condorcet's Model and Its Refinements

In his original paper, Condorcet made the following assumptions.

(1) In every pairwise comparison each voter chooses the better alternative with some fixed probability p, where $1/2 < p \leq 1$.
(2) Each voter's judgment on every pair of alternatives is independent of her judgment on every other pair.
(3) Each voter's judgment is independent of the other voters' judgments.
(4) Each voter's judgment produces a ranking of the alternatives.

However, assumptions (2) and (4) are incompatible. Indeed, if a voter ranks every pair of alternatives correctly with some fixed probability, then she may end up with a non-transitive judgment, which is prohibited by (4). In other words, if we insist that voters always produce a linear order as their judgment, then their judgments on different pairs of alternatives are no longer independent.

There are two differing opinions on how exactly Condorcet's model should be understood. Some believe that we should allow intransitive preferences, arguing that the vote is not really a preference, but rather the voter's best approximation to the correct ranking as she perceives it. It may happen that the best approximation is in fact intransitive (see, e.g., Truchon, 2008); however, it cannot be ignored, as it provides useful information.

Another interpretation of Condorcet's proposal is as follows: a voter forms her opinion by considering pairs of alternatives independently, but if the result happens to be intransitive, she discards it and tries to form her opinion again until a valid (acyclic) preference order is obtained. In statistics, the resulting probabilistic model is known as the *Mallows noise model* (Mallows, 1957). Note, however, that this model violates condition (2) (see, e.g., Gordon and Truchon, 2008).

Commenting on Condorcet's writings, Young (1988) wrote: "One must admit that the specific probabilistic model by which Condorcet reached his conclusions is almost certainly not correct in its details." He went further to say that the plausibility of any solution based on Condorcet's ideas must therefore be subjected to other tests. However, he went on and developed Condorcet's framework to see what Condorcet would have obtained if he possessed the necessary technical skills to perform his analysis to the end. We will now present Young's analysis, together with some refinements and extensions.

8.3.3 MLE for Choosing a Ranking

In this section, we describe an MLE approach to selecting the best ranking(s) of the alternatives. Recall that a *social preference function* is a mapping that given a list of rankings of the alternatives outputs a non-empty set of aggregate rankings; thus, in this section we focus on representing social preference functions within the MLE framework.

We start by presenting Young's analysis of Condorcet's proposal (see Young, 1988), followed by a discussion of a more general approach put forward by Conitzer and Sandholm (2005) and Conitzer et al. (2009).

Let $u \in \mathcal{L}(A)$ be the true state of the world, and let $v \in \mathcal{L}(A)$ be some ranking that agrees with u on k pairs of alternatives. Note that we have $d_{\text{swap}}(v,u) = \binom{m}{2} - k$. Then under both interpretations of Condorcet's model discussed in Section 8.3.2 the probability that a voter forms opinion v is proportional to

$$p^k(1-p)^{\binom{m}{2}-k} = p^{\binom{m}{2}-d_{\text{swap}}(v,u)}(1-p)^{d_{\text{swap}}(v,u)}.$$

If each voter forms her opinion independently from other voters, the probability of a profile $(v_1, \ldots, v_n)$ given that u is the true state of the world is proportional to

$$\prod_{i=1}^{n} \left(\frac{p}{1-p}\right)^{-d_{\text{swap}}(v_i,u)} = \left(\frac{p}{1-p}\right)^{-\sum_{i=1}^{n} d_{\text{swap}}(v_i,u)}.$$

If each state of the world is *a priori* considered equally likely, the rankings that are most likely to be correct are the ones that maximize the probability of the observed data, or, equivalently, minimize $\sum_{i=1}^{n} d_{\text{swap}}(v_i, u)$ (note that $p > 1/2$ and hence $\frac{p}{1-p} > 1$). Thus, Condorcet's approach results in a social preference function f_{Cond} that given a profile $R = (v_1, \ldots, v_n)$ over a candidate set A, outputs the set $\operatorname{argmin}_{u \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{\text{swap}}(v_i, u)$. This is exactly the social preference function associated with the Kemeny rule (see Section 8.2.1).

General r-Noise Models

Young's analysis is based on a specific *noise model*, i.e., a way voters' judgments are formed given an underlying state of the world. By considering other noise models, we can obtain other social preference functions. To pursue this agenda, we need a formal definition of a noise model.

Definition 8.16 A *noise model for rankings*, or an *r-noise model*, over a candidate set A is a family of probability distributions $\mathcal{P}(\cdot \mid u)_{u \in \mathcal{L}(A)}$ on $\mathcal{L}(A)$. For a given $u \in \mathcal{L}(A)$, $\mathcal{P}(v \mid u)$ is the probability that a voter forms a preference order v when the correct ranking is u.

We emphasize that the parameters of a noise model are assumed to be the same for all voters and do not depend on the number of voters. That is, we think of voters as independent agents that are influenced by the same factors in the same way.

Example 8.17 The *Mallows model* (Mallows, 1957) is a family of r-noise models $\big(\mathcal{P}_{d_{\mathrm{swap}},p}\big)_{1/2<p<1}$ given by

$$\mathcal{P}_{d_{\mathrm{swap}},p}(v \mid u) = \frac{1}{\mu_p}\varphi^{-d_{\mathrm{swap}}(v,u)}, \quad \text{where} \quad \varphi = \frac{p}{1-p} \quad \text{and} \quad \mu_p = \sum_{v \in \mathcal{L}(A)} \varphi^{-d_{\mathrm{swap}}(v,u)}.$$

Here, μ_p is the normalization constant; since d_{swap} is a neutral distance, the value of μ_p does not depend on the choice of u (Mallows, 1957).

Under the MLE approach, every r-noise model leads to a social preference function.

Definition 8.18 A social preference function f over A is the *maximum likelihood estimator (MLE) for an r-noise model $\mathcal{P}$ over A* if for every positive integer n and every n-voter profile $R = (v_1, \ldots, v_n)$ it holds that

$$f(R) = \underset{u \in \mathcal{L}(A)}{\operatorname{argmax}} \prod_{i=1}^{n} \mathcal{P}(v_i \mid u).$$

A very general method of constructing r-noise models was proposed by Conitzer et al. (2009), who introduced the notion of a *simple ranking scoring function*.

Definition 8.19 A social preference function f over A is said to be a *simple ranking scoring function (SRSF)* if there exists a mapping $\rho : \mathcal{L}(A) \times \mathcal{L}(A) \to \mathbb{R}$ such that for every positive integer n and every n-voter profile $R = (v_1, \ldots, v_n)$ it holds that

$$f(R) = \underset{u \in \mathcal{L}(A)}{\operatorname{argmax}} \sum_{i=1}^{n} \rho(v_i, u). \tag{8.5}$$

Intuitively, $\rho(v, u)$ assigns a score to v based on the similarity between v and u, and f chooses u so as to maximize the total score of the given profile. We say that a mapping $\rho : \mathcal{L}(A) \times \mathcal{L}(A) \to \mathbb{R}$ is *neutral* if $\rho(v', u') = \rho(v, u)$, where rankings v' and u' are obtained by renaming alternatives in v and u according to some permutation $\pi : A \to A$. Conitzer et al. (2009) show that a simple ranking scoring function f is neutral if and only if there exists a neutral mapping ρ satisfying (8.5).

Example 8.20 Every distance d on $\mathcal{L}(A)$ defines a simple ranking scoring function: we can set $\rho(v, u) = -d(v, u)$. The corresponding social preference function maps a profile $R = (v_1, \dots, v_n)$ to the set of rankings $\text{argmin}_{u \in \mathcal{L}(A)} \sum_{i=1}^{n} d(v_i, u)$. Observe that this social preference function is closely related to the voting rule that is distance rationalizable via $\widehat{d}$ and the strong unanimity consensus $\mathcal{S}$.

Every SRSF corresponds to an infinite family of r-noise models: If f is the SRSF defined by a mapping ρ, then for every $\varphi \in (1, +\infty)$ we can set

$$\mathcal{P}_{\rho,\varphi}(v \mid u) = \frac{1}{\mu_{\rho,\varphi,u}} \varphi^{\rho(v,u)}, \quad \text{where} \quad \mu_{\rho,\varphi,u} = \sum_{v \in \mathcal{L}(A)} \varphi^{\rho(v,u)}; \tag{8.6}$$

Conitzer et al. (2009) use $\varphi = 2$ in their paper. By construction, f is the maximum likelihood estimator for $\mathcal{P}_{\rho,\varphi}$ for every $\varphi \in (1, +\infty)$.

Conitzer et al. (2009) show that for social preference functions that are neutral (i.e., their output does not depend on the names of the candidates) the converse is also true. More precisely, they prove the following characterization result.

Theorem 8.21 *A neutral social preference function is an MLE if and only if it is an SRSF.*

Theorem 8.21 provides a convenient way to show that a given social preference function f is an MLE: it suffices to exhibit a mapping ρ witnessing that f is an SRSF. Conitzer et al. (2009) apply this method to show that for every score vector $\alpha = (\alpha_1, \dots, \alpha_m)$ the corresponding social preference function f_α is an MLE. In the rest of this section, we give a sketch of their argument.

Recall that f_α is the social preference function that orders the candidates by their α-scores, where the α-score of a candidate a in a profile $R = (v_1, \dots, v_n)$ is given by $s_\alpha(R, a) = \sum_{i=1}^{n} \alpha_{\text{pos}(v_i,a)}$; if some candidates have the same score, f_α outputs all rankings that can be obtained by breaking such ties in some way.

To show that f_α is an SRSF, let $\beta_1, \dots, \beta_m$ be a monotonically decreasing sequence (e.g., we can take $\beta_j = m - j$), and set

$$\rho_\alpha(v, u) = \sum_{a \in A} \beta_{\text{pos}(u,a)} \alpha_{\text{pos}(v,a)}, \tag{8.7}$$

We claim that f_α is the simple ranking scoring function that corresponds to ρ_α. Indeed, for a given profile $R = (v_1, \dots, v_n)$ we obtain

$$\sum_{i=1}^{n} \rho_\alpha(v_i, u) = \sum_{a \in A} \beta_{\text{pos}(u,a)} \left(\sum_{i=1}^{n} \alpha_{\text{pos}(v_i,a)} \right) = \sum_{a \in A} \beta_{\text{pos}(u,a)} s_\alpha(R, a).$$

Thus, for u to maximize the expression $\sum_{i=1}^{n} \rho_\alpha(v_i, u)$, we should have $\beta_{\mathrm{pos}(u,a)} > \beta_{\mathrm{pos}(u,b)}$ (and hence $\mathrm{pos}(u, a) < \mathrm{pos}(u, b)$) whenever $s_\alpha(R, a) > s_\alpha(R, b)$, i.e., u orders the candidates by their α-score from the highest to the lowest, breaking ties arbitrarily. Theorem 8.21 then implies the following corollary.

Corollary 8.22 *The social preference function f_α is an MLE.*

8.3.4 MLE for Choosing a Winner

In the previous section we described an MLE approach to selecting the best ranking(s). However, typically our goal is to select a single winner (or possibly a set of winners) rather than a ranking of the candidates. To extend the MLE framework to this setting, we can simply output the top candidate(s) in the best ranking(s). Alternatively, we can estimate the likelihood that a given candidate is the best. To this end, for each candidate we determine the total probability mass (with respect to the uniform distribution) of the rankings where she is the top choice, and output the candidate(s) that maximize this quantity; the validity of this method follows from the Bayes rule. We will now discuss these approaches in more detail.

Deducing Winners from Rankings: MLERIV Rules

In Section 8.2.1 we transformed the social preference function associated with the Kemeny rule into a voting rule, by picking the top candidate in each ranking output by this social preference function. By extending this procedure to arbitrary MLE social preference functions, we obtain a class of rules known as MLERIV (Conitzer and Sandholm, 2005).

Definition 8.23 Let f be a social preference function that is MLE for an r-noise model $\mathcal{P}$. Let $\hat{f}$ be a voting rule defined by $\hat{f}(R) = \{\mathrm{top}(u) \mid u \in f(R)\}$, where $\mathrm{top}(u)$ denotes the top candidate in ranking u. This rule is called the *maximum likelihood estimator for ranking under identically distributed independent votes (MLERIV) for $\mathcal{P}$.*

According to Definition 8.23, the Kemeny rule is MLERIV for the Mallows noise model. Another family of MLERIV rules is provided by Example 8.20: Theorem 8.21 implies that for every neutral distance d over $\mathcal{L}(A)$ the $(\mathcal{S}, \widehat{d})$-rationalizable voting rule is MLERIV. Further, Corollary 8.22 implies that every scoring rule is MLERIV.

Estimating the Winners: Young's Interpretation of Condorcet's Proposal

The MLERIV-based approach provides a simple way to cast many voting rules within the MLE framework. However, it is not appropriate if our goal is to output the candidate that is most likely to be ranked first. Indeed, under an r-noise model the probability that a candidate is ranked first in the true ranking is obtained by adding together the probabilities of all rankings where she appears on top, and it is entirely possible that the top candidate in the most likely ranking is a, but the cumulative probability of rankings that have a on top is lower than the cumulative probability of rankings that have some other candidate b on top.

This was clearly understood by Condorcet himself, who probably did not have the technical skills to pursue this line of reasoning. Young (1988) argues that this approach would lead him to the Borda rule, at least when p is sufficiently close to $1/2$. Young also speculates on reasons why Condorcet might have chosen to abandon this train of thought (see Young, 1988, for an amusing account of the relationship between Condorcet and Borda).

We will now present Young's extension of Condorcet's analysis. While it aims to estimate the most likely winner under the Mallows model, it makes the simplifying assumption that in the prior distribution over the states of the world all pairwise comparisons between the alternatives are independent from each other. For the Mallows model this assumption is not true: if $A = \{a, b, c\}$ and the prior distribution over the states of the world is uniform over $\mathcal{L}(A)$, knowing that in the true state of the world a is ranked above b influences our beliefs about the outcome of the comparison between a and c. Thus, Young's analysis can be seen as a heuristic algorithm for computing the most likely winner; later, we will see that its output may differ from that of the exact algorithm (see also Xia, 2014a).

Given a pair of candidates $a, b \in A$, let n_{ab} denote the number of voters in a given profile $(v_1 \dots, v_n)$ who prefer a to b. Let S be a fixed set of voters of size n_{ab}, and consider the event that the voters in S prefer a to b, while the remaining voters prefer b to a; denote this event by $\mathcal{E}_S$. If in the true state of the world a is preferred to b, then the probability of $\mathcal{E}_S$ is exactly $p^{n_{ab}}(1-p)^{n_{ba}}$. Conversely, if in the true state of the world b is preferred to a, then the probability of $\mathcal{E}_S$ is $(1-p)^{n_{ab}}p^{n_{ba}}$. The prior probability that in the true state of the world a is preferred b is exactly $1/2$. Therefore, the probability of the event $\mathcal{E}_S$ is $\frac{1}{2}(p^{n_{ab}}(1-p)^{n_{ba}} + (1-p)^{n_{ab}}p^{n_{ba}})$. Hence, by the Bayes rule, the probability that in the true state of the world a is preferred to b is proportional to

$$\frac{p^{n_{ab}}(1-p)^{n_{ba}}}{p^{n_{ab}}(1-p)^{n_{ba}} + (1-p)^{n_{ab}}p^{n_{ba}}}. \tag{8.8}$$

To compute the probability that in the true state of the world a is preferred to every other candidate, we take the product of probabilities (8.8) over all $b \neq a$; note that this step makes use of the assumption that in the prior distribution over the states of the world all pairwise comparisons are independent. It follows that the probability that a is the true winner given that the observed profile is $(v_1, \dots, v_n)$ is given by

$$\prod_{b \in A \setminus \{a\}} \frac{p^{n_{ab}}(1-p)^{n_{ba}}}{p^{n_{ab}}(1-p)^{n_{ba}} + (1-p)^{n_{ab}}p^{n_{ba}}} = \prod_{b \in A \setminus \{a\}} \frac{1}{1 + \left(\frac{1-p}{p}\right)^{n_{ab}-n_{ba}}}.$$

Thus, the most likely winners are the candidates that minimize the expression

$$\kappa_a(\varphi) = \prod_{b \in A \setminus \{a\}} \left(1 + \varphi^{n_{ba} - n_{ab}}\right), \quad \text{where} \quad \varphi = \frac{p}{1-p}. \tag{8.9}$$

Now, the behavior of this expression crucially depends on the value of $\varphi = \frac{p}{1-p}$. We will consider two cases: (1) p is very close to 1 and hence $\varphi \to +\infty$ (i.e., a voter is almost always right) and (2) p is very close to $1/2$ and hence $\varphi \to 1$ (a voter has only a slight advantage over a random coin toss). We denote the corresponding voting rules by $\text{MLE}^{\infty}_{\text{intr}}$ and $\text{MLE}^{1}_{\text{intr}}$, respectively (the reasons for this notation are explained in Remark 8.24). The analysis below is based on the work of Elkind and Shah (2014).

$p \to 1, \varphi \to +\infty$ The rate of growth of $\kappa_a(\varphi)$ as $\varphi \to +\infty$ depends on the degree of its highest-order term, i.e., $\sum_{b \in A \setminus \{a\} : n_{ba} > n_{ab}} (n_{ba} - n_{ab})$: slowest-growing functions correspond to the most likely candidates.

Thus, to determine the $\text{MLE}^{\infty}_{\text{intr}}$-winners, we first compute the score of each candidate $a \in A$ as the sum of a's loss margins in all pairwise elections she loses: $s_T(a) = \sum_{b \in A \setminus \{a\} : n_{ba} > n_{ab}} (n_{ba} - n_{ab})$. If there is a unique candidate with the minimum score, this candidate wins. In case of a tie among $a_1, \dots, a_k$, $\text{MLE}^{\infty}_{\text{intr}}$ takes into account the coefficients of the highest-order terms as well as the lower-order terms of $\kappa_{a_1}(\varphi), \dots, \kappa_{a_k}(\varphi)$; the resulting tie-breaking procedure is quite complicated (but can be shown to be polynomial-time computable). The voting rule that outputs the set $\arg\min_{a \in A} s_T(a)$ was proposed by Tideman (1987) as an approximation to the Dodgson rule, and is now known as the Tideman rule; thus, our analysis shows that $\text{MLE}^{\infty}_{\text{intr}}$ is a refinement of the Tideman rule. The Tideman rule has been studied by McCabe-Dansted et al. (2008), as well as by Caragiannis et al. (2014), who refer to it as the simplified Dodgson rule; an overview of their results can be found in Chapter 5 (Caragiannis et al., 2015)[4].

$p \to 1/2, \varphi \to 1$ In this case, we are interested in the behavior of $\kappa_a(\varphi)$ as $\varphi \to 1$. We have $\kappa_a(1) = 2^{m-1}$ for all $a \in A$. Further, the derivative of $\kappa_a(\varphi)$ at $\varphi = 1$ is $\sum_{c \neq a}(n_{ca} - n_{ac})2^{m-2} = \sum_{c \neq a}(n - 2n_{ac})2^{m-2}$. To minimize this expression, we need to maximize $\sum_{c \neq a} n_{ac}$, which is the Borda score of a. Hence, $\text{MLE}^{1}_{\text{intr}}$ is a refinement of the Borda rule: it selects the Borda winner when it is unique, and if there are several Borda winners, it breaks ties by taking into account higher-order derivatives of $\kappa_a(\varphi)$ at $\varphi \to 1$.

Remark 8.24 One can think of Young's procedure as estimating the most likely winner under a different noise model, namely, one where the prior distribution assigns equal probability to all tournaments over A, i.e., the state of the world is described by the outcomes of $\binom{m}{2}$ comparisons, and all vectors of outcomes are

[4] Young (1988) appears to suggest that $\text{MLE}^{\infty}_{\text{intr}}$ is Maximin; our argument shows that this is not the case.

considered to be equally likely. Voters' preferences are tournaments as well; in each vote, the direction of every edge agrees with the ground truth with probability p and disagrees with it with probability $1-p$, with decisions for different edges made independently from each other. We emphasize that this distribution assigns non-zero probability to "states of the world" that violate transitivity. For this noise model, Young's procedure correctly identifies the candidate with the largest cumulative probability of the states of the world where she wins all her pairwise elections.

It is often claimed that $\mathrm{MLE}^1_{\mathrm{intr}}$ is the Borda rule. We will now show that this claim is inaccurate: while $\mathrm{MLE}^1_{\mathrm{intr}}$ chooses among the Borda winners, it may fail to select some of them.

Example 8.25 Let $A = \{a, b, c, d\}$ and consider a 4-voter profile over A given by $(adcb, bcad, abdc, bcad)$ (where we write $xyzt$ as a shorthand for $x \succ y \succ z \succ t$). The Borda winners in this profile are a and b, and their Borda score is 8. On the other hand, we have $\kappa_a(\varphi) = 4(1+\varphi^{-4})$, $\kappa_b(\varphi) = 2(1+\varphi^{-2})^2$. The reader can verify that $\kappa_a(1.2) \approx 5.93$, $\kappa_b(1.2) \approx 5.74$, and

$$\frac{d\kappa_a}{d\varphi}\Big|_{\varphi=1} = \frac{d\kappa_b}{d\varphi}\Big|_{\varphi=1} = -16, \quad \text{but} \quad \frac{d^2\kappa_a}{(d\varphi)^2}\Big|_{\varphi=1} = 48, \quad \frac{d^2\kappa_b}{(d\varphi)^2}\Big|_{\varphi=1} = 32,$$

so b emerges as the unique winner under $\mathrm{MLE}^1_{\mathrm{intr}}$.

Estimating the Winners Under an r-noise Model

It is natural to ask whether we can estimate the most likely winner under the Mallows model without making the simplifying assumption that in the prior distribution over the states of the world all pairwise comparisons are independent. To the best of our knowledge, Procaccia et al. (2012) were the first to do this for $p \to 1/2$; their argument extends to more general noise models and to settings where the goal is to select a fixed-size subset of candidates. They have also considered the case $p \to 1$ (see also the work of Elkind and Shah, 2014). Just as in Young's analysis, the result turns out to depend on the value of p: when $p \to 1$ (and $\varphi = \frac{p}{1-p} \to +\infty$), we obtain a refinement of the Kemeny rule, and when $p \to 1/2$ (and $\varphi \to 1$), we obtain a refinement of the Borda rule. We will now present the arguments both for $\varphi \to +\infty$ and for $\varphi \to 1$; we refer to the resulting rules as $\mathrm{MLE}^\infty_{\mathrm{tr}}$ and $\mathrm{MLE}^1_{\mathrm{tr}}$, respectively.

For every candidate $a \in A$ let $\mathcal{L}_a$ denote the set of all rankings in $\mathcal{L}(A)$ where a is ranked first. Recall that under the Mallows noise model the probability of a profile $(v_1, \dots, v_n)$ given that the true state of the world is described by a ranking u is proportional to $\varphi^{-\sum_{i=1}^n d_{\mathrm{swap}}(v_i, u)}$. Thus, to compute the most likely winner, we need to find the candidates that maximize the expression

$$\tau_a(\varphi) = \sum_{u \in \mathcal{L}_a} \varphi^{-\sum_{i=1}^n d_{\mathrm{swap}}(v_i, u)}.$$

$\boldsymbol{p \to 1, \varphi \to +\infty}$ The rule $\mathrm{MLE}^\infty_{\mathrm{tr}}$ returns a set of candidates S such that for every $a \in S$, $b \in A \setminus S$ we have $\tau_a(\varphi) > \tau_b(\varphi)$ for all sufficiently large values of φ. To see that S is not empty, note that functions $\tau_a(\varphi)$, $a \in A$, are Laurent polynomials (i.e., sums of powers of φ), and therefore any two of these functions either coincide or have finitely many intersection points. Moreover, for each $a \in A$ the most significant summand of $\tau_a(\varphi)$ at $\varphi \to +\infty$ is

$$\varphi^{-\sum_{i=1}^{n} d_{\text{swap}}(v_i,u')}, \quad \text{where} \quad u' \in \operatorname{argmin}_{u \in \mathcal{L}_a} \sum_{i=1}^{n} d_{\text{swap}}(v_i, u).$$

Hence, $\text{MLE}_{\text{tr}}^{\infty}$ is a refinement of the Kemeny rule.

$p \to 1/2, \varphi \to 1$ We have $\tau_a(1) = (m-1)!$ for all $a \in A$. Further, the derivative of $\tau_a(\varphi)$ at $\varphi = 1$ is given by

$$\frac{d\tau_a}{d\varphi}\Big|_{\varphi=1} = -\sum_{u \in \mathcal{L}_a} \sum_{i=1}^{n} d_{\text{swap}}(v_i, u) = -\sum_{i=1}^{n} \sum_{u \in \mathcal{L}_a} d_{\text{swap}}(v_i, u).$$

It is easy to show by induction on j that if $\text{pos}(v_i, a) = j$ then we have $\sum_{u \in \mathcal{L}_a} d_{\text{swap}}(v_i, u) = j(m-1)! + C_m$, where C_m is a function of m (i.e., does not depend on v_i). As $\sum_{i=1}^{n}(m - \text{pos}(v_i, a))$ is exactly the Borda score of a, it follows that $a \in \operatorname{argmin}_{c \in A} \frac{d\tau_c}{d\varphi}\Big|_{\varphi=1}$ if and only if a is a Borda winner. Hence, $\text{MLE}_{\text{tr}}^{1}$ is a refinement of the Borda rule. Further, it can be checked that it is distinct from the Borda rule, i.e., it may fail to elect some Borda winners; this can happen when $\tau_a(\varphi)$ and $\tau_b(\varphi)$ are different from each other, even though their derivatives at $\varphi = 1$ coincide. Further, it can also be shown that $\text{MLE}_{\text{tr}}^{1} \neq \text{MLE}_{\text{intr}}^{1}$ (Elkind and Shah, 2014), i.e., these two rules are two distinct refinements of the Borda rule.

We can apply a similar procedure to other r-noise models. It turns out that for noise models that are derived from neutral simple ranking scoring functions via equation (8.6) in the case $\varphi \to 1$ we obtain a voting rule that is a refinement of some scoring rule.

In more detail, consider a neutral SRSF given by a mapping $\rho : \mathcal{L}(A) \times \mathcal{L}(A) \to \mathbb{R}$, a value $\varphi \in (1, +\infty)$, and the corresponding r-noise model $\mathcal{P}_{\rho,\varphi}(v \mid u) = \frac{1}{\mu_{\rho,\phi,u}} \varphi^{\rho(v,u)}$. Since ρ is neutral, $\mu_{\rho,\varphi,u}$ is the same for all $u \in \mathcal{L}(A)$. Assume that each ranking of the aletrnatives is *a priori* equally likely. A direct application of the Bayes rule shows that the probability that the true state of the world is a ranking where a is placed first given that the input profile is $(v_1, \dots, v_n)$ is proportional to

$$\sum_{u \in \mathcal{L}_a} \varphi^{\sum_{i=1}^{n} \rho(v_i,u)}. \tag{8.10}$$

Let MLE_{ρ}^{1} be the voting rule that maps $(v_1, \dots, v_n)$ to the set of candidates that maximize expression (8.10) for values of φ that are close to 1.

We can view expression (8.10) as a function of φ; its derivative at $\varphi = 1$ equals

$$\sum_{u \in \mathcal{L}_a} \sum_{i=1}^{n} \rho(v_i, u) = \sum_{i=1}^{n} \sum_{u \in \mathcal{L}_a} \rho(v_i, u).$$

This means that the set of MLE^1_ρ-winners is a (possibly strict) subset of $W = \mathrm{argmax}_{a \in A} \sum_{i=1}^{n} \sum_{u \in \mathcal{L}_a} \rho(v_i, u)$. Let $\overline{\mathrm{MLE}^1_\rho}$ be a coarsening of MLE^1_ρ that, given a profile $(v_1, \ldots, v_n)$, outputs the entire set W. Since ρ is neutral, the value of the expression $\sum_{u \in \mathcal{L}_a} \rho(v_i, u)$ only depends on the position of a in v_i. Thus, $\overline{\mathrm{MLE}^1_\rho}$ is a scoring rule. Conversely, every scoring rule can be obtained as $\overline{\mathrm{MLE}^1_\rho}$ for a suitable function ρ: e.g., for the rule f_α we can use the function ρ_α defined by (8.7).

Noise Models for Winners: MLEVIW Rules

We have seen how to derive a voting rule from an r-noise model by considering the cumulative probability of rankings with a given winner. Conitzer and Sandholm (2005) put forward a direct MLE-based approach for defining voting rules. It is based on a simplified noise model, where the "state of the world" is simply the identity of the best candidate, and the likelihood of a given vote depends on the position of this candidate in the vote.

Definition 8.26 A *noise model for winners*, or a *w-noise model*, over a candidate set A, $|A| = m$, is a family of probability distributions $\overline{\mathcal{P}}(\cdot \mid a)_{a \in A}$ on $\{1, \ldots, m\}$. For a given $a \in A$, $\overline{\mathcal{P}}(j \mid a)$ is the probability of a vote where a is ranked in position j given that a is the correct winner. We require $\overline{\mathcal{P}}(j \mid a) > 0$ for all $a \in A$, $j = 1, \ldots, m$.

A voting rule f over a candidate set A is a *maximum likelihood estimator for winner under identically distributed independent votes (MLEWIV)* with respect to a w-noise model $\overline{\mathcal{P}}$ over A if for every positive integer n and every preference profile $R = (v_1, \ldots, v_n) \in \mathcal{L}(A)^n$ it holds that

$$f(R) = \operatorname*{argmax}_{a \in A} \prod_{i=1}^{n} \overline{\mathcal{P}}(\mathrm{pos}(v_i, a) \mid a). \qquad (8.11)$$

However, the power of this approach is somewhat limited, at least if we require neutrality: neutral MLEWIV rules are simply scoring rules (Conitzer and Sandholm, 2005; Elkind et al., 2010b). Note the some form of neutrality is implicit in the definition of a w-noise model: by construction, this model assigns the same probability to any two votes that rank a in the same position, irrespective of how they rank the remaining candidates.

Proposition 8.27 *For every score vector $\alpha = (\alpha_1, \ldots, \alpha_m)$ the scoring rule f_α is MLEWIV. Conversely, every neutral MLEWIV rule is a scoring rule.*

Proof Given a score vector $\alpha = (\alpha_1, \ldots, \alpha_m)$, define a w-noise model $\overline{\mathcal{P}}_\alpha$ as

$\overline{\mathcal{P}}_\alpha(j \mid a) = \frac{1}{\mu_\alpha} 2^{\alpha_j}$, where $\mu_\alpha = \sum_{j=1}^{m} 2^{\alpha_j}$. Now, consider an arbitrary profile $R = (v_1, \ldots, v_n)$ over A and a candidate $a \in A$. For each $i = 1, \ldots, n$, let $p_i = \mathrm{pos}(v_i, a)$. The α-score of a in R is given by $s_\alpha(R, a) = \sum_{i=1}^{n} \alpha_{p_i}$. On the other hand, we have

$$\prod_{i=1}^{n} \overline{\mathcal{P}}_\alpha(\mathrm{pos}(v_i, a) \mid a) = \frac{1}{\mu_\alpha^n} \prod_{i=1}^{n} 2^{\alpha_{p_i}} = \frac{1}{\mu_\alpha^n} 2^{s_\alpha(a,R)}. \tag{8.12}$$

Hence, the set of most likely candidates under $\overline{\mathcal{P}}_\alpha$ is exactly the set of f_α-winners.

Conversely, let f be a neutral MLEWIV rule for a w-noise model $\overline{\mathcal{P}}$. It is easy to verify that $\overline{\mathcal{P}}$ is neutral, i.e., $\overline{\mathcal{P}}(j \mid a) = \overline{\mathcal{P}}(j \mid b)$ for every $j = 1, \ldots, m$ and every $a, b \in A$. Now, fix some $a \in A$ and set $\alpha_j = \log_2 \overline{\mathcal{P}}(j \mid a)$ for all $j = 1, \ldots, m$. Equation (8.12) shows that the scoring rule f_α coincides with f. □

Proposition 8.27 provides an alternative characterization of scoring rules, thus complementing the well-known results of Smith (1973) and Young (1975). Equivalently, one can say that the results of Smith and Young provide a characterization of MLEWIV rules in terms of standard axiomatic properties. A natural open question, which was suggested by Conitzer et al. (2009), is whether a similar characterization can be obtained for MLERIV rules.

To conclude our discussion of the MLEWIV rules, we note that these rules arise naturally from the ranking-based model considered in the previous section. Indeed, for a neutral function ρ the rule $\overline{\mathrm{MLE}}_\rho^1$ is MLEWIV. To see this, note that given a candidate $a \in A$, we can pick $\varphi \in (1, +\infty)$ and m rankings $v^1, \ldots, v^m$ such that $\mathrm{pos}(v^j, a) = j$ for $j = 1, \ldots, m$, and set

$$\overline{\mathcal{P}}(j \mid a) = \frac{1}{\mu} \varphi^{\sum_{u \in \mathcal{L}_a} \rho(v^j, u)}, \quad \text{where} \quad \mu = \sum_{j=1}^{m} \varphi^{\sum_{u \in \mathcal{L}_a} \rho(v^j, u)}.$$

It is easy to verify that for any choice of $\varphi \in (1, +\infty)$ and $v^1, \ldots, v^m$ the MLEWIV rule that corresponds to this noise model is exactly $\overline{\mathrm{MLE}}_\rho^1$.

Finally, we remark that Ben-Yashar and Paroush (2001) consider another approach to estimating winners under noise: in their model, each voter has to specify one candidate (rather than a ranking of the candidates), and a voter's probability of voting for the true winner depends on the identity of the winner, and may vary from one voter to another. Ben-Yashar and Paroush present an extension of Condorcet's Jury Theorem (see Section 8.3.1) to this setting.

8.4 Conclusions and Further Reading

We have discussed two approaches to rationalizing voting rules: a consensus-based approach that leads to the distance rationalizability framework and a probabilistic approach that leads to the MLE framework. We showed how to rationalize many common voting rules in each of these frameworks. For some rules, such as the Kemeny rule, the rationalizations provided by both frameworks are closely related, while for others (e.g., scoring rules), they seem to be quite different, and thus provide different perspectives on the rule in question.

Due to space constraints, we were not able to overview the entire body of research on these two frameworks; we will now briefly mention some of the relevant papers.

Service and Adams (2012) consider randomized strategyproof approximations to distance rationalizable voting rules. Boutilier and Procaccia (2012) relate the concept of distance rationalizability to the framework of dynamic social choice (Parkes and Procaccia, 2013). Distance-based approaches have also been considered in the context of judgment aggregation (Lang et al., 2011; Dietrich, 2014), as well as in other areas of social choice (see Eckert and Klamler, 2011, and references therein).

Xia et al. (2010) apply the MLE framework to voting in multi-issue domains, and Xia and Conitzer (2011) extend it to partial orders, and a more general notion of "state of the world"; for instance, they consider settings where the goal is to estimate the top k alternatives for $k \geq 1$. The latter problem is explored in more detail by Procaccia et al. (2012). Caragiannis et al. (2013b) investigate a complementary issue: given a noise model and a fixed voting rule, how many samples do we need to generate so that this rule identifies the correct winner? They also consider voting rules that perform well with respect to *families* of noise models; such rules are further explored by Caragiannis et al. (2013a) and Xia (2014b). Drissi-Bakhkhat and Truchon (2004) modify the Mallows model by relaxing the assumption that the probability of correctly ordering two alternatives is the same for all pairs of alternatives. They let this probability increase with the distance between the two alternatives in the true order, to reflect the intuition that a judge or voter is more prone to errors when confronted with two comparable alternatives than when confronted with a good alternative and a bad one. Truchon (2008) shows that when this probability increases exponentially with the distance, the resulting ranking orders the candidates according to their Borda scores. MLE analysis admits a Bayesian interpretation: if we assume the uniform prior over the true states of the world, then an MLE rule outputs the maximum *a posteriori* estimate. Pivato (2012) considers a more general class of statistical estimators (in particular, settings where the prior distribution over the possible states of the world need not be uniform) and domains other than preference aggregation (including judgment aggregation and committee selection).

Acknowledgements We would like to thank Shmuel Nitzan, Nisarg Shah and the editors of this volume for their very useful feedback, which was instrumental in improving this chapter. We are grateful to Piotr Faliszewski for his permission to use material from our joint papers.

References

Edith Elkind[5] and Arkadii Slinko[6]

Alon, N., and Spencer, J. 2008. *Probabilistic Method, 3rd edition*. Wiley.

Arrow, K. 1951. *Social Choice and Individual Values*. John Wiley and Sons.

Austen-Banks, D., and Smith, J. S. 1994. Information aggregation, rationality, and the Condorcet Jury Theorem. *The American Political Science Review*, **90**(1), 34–45.

Baharad, E., Koppel, M., Goldberger, J., and Nitzan, S. 2011. Distilling the wisdom of crowds: weighted aggregation of decisions on multiple issues. *Journal of Autonomous Agents and Mult-Agent Systems*, **22**, 31–42.

Baharad, E., Koppel, M., Goldberger, J., and Nitzan, S. 2012. Beyond Condorcet: optimal judgment aggregation using voting records. *Theory and Decision*, **72**, 113–130.

Baigent, N. 1987. Metric rationalisation of social choice functions according to principles of social choice. *Mathematical Social Sciences*, **13**(1), 59–65.

Balinski, L., and Laraki, R. 2010. *Majority Judgment: Measuring, Ranking, and Electing*. MIT Press.

Ben-Yashar, R. C., and Nitzan, S. 1997. The optimal decision rule for fixed-size committees in dichotomous choice situations: the general result. *International Economic Review*, **38**(1), 175–186.

Ben-Yashar, R. C., and Paroush, J. 2001. Optimal decision rules for fixed-size committees in polychotomous choice situations. *Social Choice and Welfare*, **18**(4), 737–746.

Berg, S. 1993a. Condorcet's jury theorem, dependency among jurors. *Social Choice and Welfare*, **10**(1), 87–95.

Berg, S. 1993b. Condorcet's jury theorem revisited. *European Journal of Political Economy*, **9**(3), 437–446.

Bossert, W., and Suzumura, K. 2010. *Consistency, Choice, and Rationality*. Harvard University Press.

Boutilier, C., and Procaccia, A. 2012. A dynamic rationalization of distance rationalizability. Pages 1278–1284 of: *AAAI'12*.

Brams, S., and Fishburn, P. 2002. Voting procedures. Pages 173–236 of: Arrow, K., Sen, A., and Suzumura, K. (eds), *Handbook of Social Choice and Welfare, Volume 1*. Elsevier.

Campbell, D., and Nitzan, S. 1986. Social compromise and social metrics. *Social Choice and Welfare*, **3**(1), 1–16.

Camps, R., Mora, X., and Saumell, L. 2014. Social choice rules driven by propositional logic. *Annals of Mathematics and Artificial Intelligence*, **70**(3), 279–312.

Caragiannis, I., Procaccia, A., and Shah, N. 2013a. Modal ranking: A uniquely robust voting rule. Pages 616–622 of: *AAAI'14*.

Caragiannis, I., Procaccia, A., and Shah, N. 2013b. When do noisy votes reveal the truth? Pages 143–160 of: *ACM EC'13*.

Caragiannis, I., Kaklamanis, C., Karanikolas, N., and Procaccia, A. 2014. Socially desirable approximations for Dodgson's voting rule. *ACM Transactions on Algorithms*, **10**(2).

Caragiannis, Ioannis, Hemaspaandra, Edith, and Hemaspaandra, Lane A. 2015. Dodgson's Rule and Young's Rule. Chap. 5 of: Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D. (eds), *Handbook of Computational Social Choice*. Cambridge University Press.

Chebotarev, P. Yu., and Shamis, E. 1998. Characterizations of scoring methods for preference aggregation. *Annals of Operations Research*, **80**, 299–332.

Condorcet, J.-A.-N. de Caritat, Marquis de. 1785. *Essai sur l'Application de L'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Facsimile reprint of original published in Paris, 1972, by the Imprimerie Royale.

Conitzer, V., and Sandholm, T. 2005. Common voting rules as maximum likelihood estimators. Pages 145–152 of: *UAI'05*.

Conitzer, V., Rognlie, M., and Xia, L. 2009. Preference functions that score rankings and maximum likelihood estimation. Pages 109–115 of: *IJCAI'09*.

Dietrich, F. 2014. Scoring rules for judgment aggregation. *Social Choice and Welfare*, **42**, 873–911.

Dietrich, F., and List, C. 2004. A model of jury decision where all the jurors have the same evidence. *Synthese*, **142**, 175–202.

Dodgson, C. 1876. *A method of taking votes on more than two issues*. Pamphlet printed by the Clarendon Press, Oxford, and headed "not yet published".

Drissi-Bakhkhat, M., and Truchon, M. 2004. Maximum likelihood approach to vote aggregation with variable probabilities. *Social Choice and Welfare*, **23**(2), 161–185.

Eckert, D., and Klamler, C. 2011. Distance-based aggregation theory. Pages 3–22 of: *Consensual Processes*. Springer.

Elkind, E., and Shah, N. 2014. Electing the most probable without eliminating the irrational: voting over intransitive domains. Pages 182–191 of: *UAI'14*.

Elkind, E., Faliszewski, P., and Slinko, A. 2009. On distance rationalizability of some voting rules. Pages 201–214 of: *TARK'09*.

Elkind, E., Faliszewski, P., and Slinko, A. 2010a. Good rationalizations of voting rules. Pages 774–779 of: *AAAI'10*.

Elkind, E., Faliszewski, P., and Slinko, A. 2010b. On the role of distances in defining voting rules. Pages 375–382 of: *AAMAS'10*.

Elkind, E., Faliszewski, P., and Slinko, A. 2011. Homogeneity and monotonicity of distance-rationalizable voting rules. Pages 821–828 of: *AAMAS'11*.

Elkind, E., Faliszewski, P., and Slinko, A. 2012. Rationalizations of Condorcet-consistent rules via distances of Hamming type. *Social Choice and Welfare*, **4**(39), 891–905.

Farkas, D., and Nitzan, S. 1979. The Borda rule and Pareto stability: a comment. *Econometrica*, **47**(5), 1305–1306.

Gordon, S., and Truchon, M. 2008. Social choice, optimal inference and figure skating. *Social Choice and Welfare*, **30**(2), 265–284.

Hemaspaandra, E., Hemaspaandra, L., and Rothe, J. 1997. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, **44**(6), 806–825.

Hemaspaandra, E., Spakowski, H., and Vogel, J. 2005. The complexity of Kemeny elections. *Theoretical Computer Science*, **349**(3), 382–391.

Hemaspaandra, L., and Ogihara, M. 2002. *The Complexity Theory Companion.* Springer-Verlag.

Kendall, M., and Gibbons, J. 1990. *Rank Correlation Methods.* Oxford University Press.

Ladha, K. K. 1992. The Condorcet Jury Theorem, free speech and correlated votes. *The American Political Science Review*, **36**(3), 617–634.

Ladha, K. K. 1993. Condorcet's jury theorem in the light of de Finetti's theorem: majority-voting with correlated votes. *Social Choice and Welfare*, **10**(1), 69–85.

Ladha, K. K. 1995. Information pooling through majority rule: Condorcet's jury theorem with correlated votes. *Journal of Economic Behavior and Organization*, **26**(3), 353–372.

Lang, J., Pigozzi, G., Slavkovik, M., and van der Torre, L. 2011. Judgment aggregation rules based on minimization. Pages 238–246 of: *TARK'11.*

Lerer, E., and Nitzan, S. 1985. Some general results on the metric rationalization for social decision rules. *Journal of Economic Theory*, **37**(1), 191–201.

Mallows, C. L. 1957. Non-null ranking models. *Biometrika*, **44**, 114–130.

May, K. 1952. A set of independent, necessary and sufficient conditions for simple majority decision. *Econometrica*, **20**(2–3), 680–684.

McCabe-Dansted, J., Pritchard, G., and Slinko, A. 2008. Approximability of Dodgson's rule. *Social Choice and Welfare*, **2**(31), 311–330.

McLennan, A. 1998. Consequences of the Condorcet Jury Theorem for beneficial information aggregation by rational agents. *The American Political Science Review*, **92**(2), 413–419.

Meskanen, T., and Nurmi, H. 2008. Closeness counts in social choice. Pages 182–191 of: Braham, M., and Steffen, F. (eds), *Power, Freedom, and Voting.* Springer-Verlag.

Nitzan, S. 1981. Some measures of closeness to unanimity and their implications. *Theory and Decision*, **13**(2), 129–138.

Nitzan, S. 2010. *Collective Preference and Choice.* Cambridge University Press.

Nitzan, S., and Paroush, J. 1982. Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review*, **23**, 289–297.

Parkes, D., and Procaccia, A. 2013. Dynamic social choice with evolving preferences. Pages 767–773 of: *AAAI'13.*

Peleg, B., and Zamir, S. 2012. Extending the Condorcet Jury Theorem to a general dependent jury. *Social Choice and Welfare*, **1**(39), 91–125.

Pivato, M. 2012. Voting rules as statistical estimators. *Social Choice and Welfare.* Online First.

Procaccia, A. D., Reddi, S., and Shah, N. 2012. A maximum likelihood approach for selecting sets of alternatives. Pages 695–704 of: *UAI'12.*

Schulze, M. 2003. A new monotonic and clone-independent single-winner election method. *Voting Matters*, **17**, 9–19.

Service, T., and Adams, J. 2012. Strategyproof approximations of distance rationalizable voting rules. Pages 569–576 of: *AAMAS'12.*

Shapley, L., and Grofman, B. 1984. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice*, **43**, 329–343.

Smith, J. H. 1973. Aggregation of preferences with variable electorate. *Econometrica*, **41**(6), 1027–1041.

Tideman, N. 1987. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, **4**(3), 185–206.

Truchon, M. 2008. Borda and the maximum likelihood approach to vote aggregation. *Mathematical Social Sciences*, **55**(1), 96–102.

Uckelman, S. L., and Uckelman, J. 2010. Strategy and manipulation in medieval elections. Pages 15–16 of: *COMSOC'10*.

Xia, L. 2014a. *Deciphering Young's interpretation of Condorcet's model*. Manuscript.

Xia, L. 2014b. Statistical properties of social choice mechanisms. In: *COMSOC'14*.

Xia, L., and Conitzer, V. 2011. A maximum likelihood approach towards aggregating partial orders. Pages 446–451 of: *IJCAI'11*.

Xia, L., Conitzer, V., and Lang, J. 2010. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. Pages 399–408 of: *AAMAS'10*.

Young, H. P. 1975. Social choice scoring functions. *SIAM Journal on Applied Mathematics*, **28**(4), 824–838.

Young, H. P. 1988. Condorcet's theory of voting. *The American Political Science Review*, **82**(4), 1231–1244.

Young, H. P., and Levenglick, A. 1978. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics*, **35**(2), 285–300.

Zwicker, William S. 2015. Introduction to Voting Theory. Chap. 2 of: Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D. (eds), *Handbook of Computational Social Choice*. Cambridge University Press.