Voting rules as statistical estimators

Pivato, Marcus
Department of Mathematics, Trent University

13. April 2011

Voting rules as statistical estimators*

Marcus Pivato, Trent University

April 13, 2011

Abstract

We adopt an 'epistemic' interpretation of social decisions: there is an objectively correct choice, each voter receives a 'noisy signal' of the correct choice, and the social objective is to aggregate these signals to make the best possible guess about the correct choice. One epistemic method is to fix a probability model and compute the maximum likelihood estimator (MLE), maximum *a posteriori* estimator (MAP) or expected utility maximizer (EUM), given the data provided by the voters. We first show that an abstract voting rule can be interpreted as MLE or MAP if and only if it is a scoring rule. We then specialize to the case of distance-based voting rules, in particular, the use of the median rule in judgement aggregation. Finally, we show how several common 'quasiutilitarian' voting rules can be interpreted as EUM.

Let $\mathcal{S}$ be a set of possible states of nature, and let $s^* \in \mathcal{S}$ be the unknown true state. Let $\mathcal{I}$ be a collection of voters, and for all $i \in \mathcal{I}$, let v^i be a signal from voter i communicating her beliefs about the true state. *Epistemic social choice theory*[1] concerns the problem of how to aggregate the opinion profile $\{v^i\}_{i\in\mathcal{I}}$ so as to make the 'best guess' about the true value of s^*.

For example, let $\mathcal{S} = \{\pm 1\}$, and suppose $\{v^i\}_{i\in\mathcal{I}}$ are independent, identically distributed (i.i.d.), $\{\pm 1\}$-valued random variables, with $\text{Prob}\,[v^i = s^*] > \frac{1}{2}$ for all $i \in \mathcal{I}$. Let $\overline{v} := \sum_{i\in\mathcal{I}} v_i$ (so $\text{sign}(\overline{v}) \in \{\pm 1\}$ is the choice of the majority). Then the well-known *Condorcet Jury Theorem* (CJT) says that $\text{Prob}\,[\text{sign}(\overline{v}) = s^*]$ approaches 1 as $|\mathcal{I}|$ becomes large. In other words, the outcome of majority vote is likely to produce the correct answer, even when the reliability of each individual voter is barely better than a coin toss.[2] Adopting an epistemic interpretation of preference aggregation, Young (1986, 1988, 1995, 1997) showed that the Kemeny rule can be seen as the maximum-likelihood estimator (MLE) of the 'true' preference ordering over a set of candidates, while the Borda rule is the MLE of the

*This paper was written while visiting the Department of Economics at the Université de Montréal, and the Centre for Philosophy of Natural and Social Sciences at the London School of Economics. I would like to thank the UdM, CIREQ, and LSE-CPNSS for their hospitality. This research was also supported by NSERC grant #262620-2008.

[1] This terminology originates with Cohen (1986) and Estlund (1997).

[2] There is now a large literature extending the CJT to choices amongst three or more alternatives, or models where voters have different competencies and/or have correlated errors. For example, see Nitzan (2010, Part III), Hummel (2010), Kaniovski (2010), or Dietrich and Spiekerman (2011).

best candidate. More recently, Conitzer and Sandholm (2005), Conitzer et al. (2009), and Xia et al. (2010) have investigated which other preference aggregators can be interpreted as MLEs.

However, preference aggregation is only one social choice problem, and not necessarily the one where the epistemic interpretation is the most plausible. This paper is concerned with the more general problem of when *any* sort of voting rule can be interpreted as a statistical 'estimator' of some kind. Section 1 introduces abstract voting rules and several kinds of of statistical estimator. Theorems 1.1 and 1.4 show that a voting rule can be interpreted as maximum likelihood or maximum *a posteriori* (MAP) estimator if and only if it is a 'scoring rule'. Examples include the Borda rule, the Kemeny rule, and approval voting. Section 2 specializes to the case when the space of social alternatives has a metric structure, which governs the sorts of errors which voters tend to make. In this setting, one MAP estimator is the *metric* voting rule, which chooses the alternative with minimal average distance to the voters (Theorem 2.1). We apply this interpretation to cyclic parameter estimation, the plurality rule, and the Borda rule. Section 3 specializes the model of §2 to the case when each voter's error probability density decays exponentially with distance from the correct solution. In this case, the metric voting rule is the *median rule.* We first apply the median rule to the estimation of a parameter on an interval (Proposition 3.2). Then we apply it to judgement aggregation (Proposition 3.3), with focus on committee selection, Arrovian preference aggregation, and certain partition problems. Finally, Section 4 considers when a voting rule can be interpreted as an expected utility maximizer (EUM). We provide EUM interpretations of approval voting, classic utilitarianism, relative utilitarianism, and variant of the Borda rule. All proofs are in an Appendix at the end of the paper.

Notation. $\mathbb{R}$ denotes the real numbers, $\mathbb{N} := \{1, 2, 3, \ldots\}$, and $\mathbb{R}_+ := [0, \infty)$. Upper-case caligraphic letters (e.g. $\mathcal{A}$, $\mathcal{I}$, $\mathcal{K}$, $\mathcal{S}$, $\mathcal{X}$, $\mathcal{V}$, etc.) will denote sets, which are either finite, or countably infinite, or assumed to be measurable subsets of nonzero Lebesgue measure in some Euclidean space $\mathbb{R}^N$. Lower-case Roman letters (e.g. a, i, k, s, x, v, etc.) will denote elements of these sets (or numbers), while upper case Roman letters generally denote functions. Boldface or sans serif letters (e.g. $\mathbf{v}$ or v) will denote n-tuples, and boldface *and* sans serif (e.g. $\boldsymbol{\mathsf{v}}$) will denote m-tuples of n-tuples. Lower-case Greek letters (e.g. α, ρ, etc.) denote functions, which are often ***probability densities*** (i.e. nonnegative functions whose total sum or integral is 1). If a set $\mathcal{X}$ is finite or countable, we always 'integrate' with respect to the counting measure on $\mathcal{X}$. Thus, if $\phi : \mathcal{X} \longrightarrow \mathbb{R}$ is any function and ρ is any probability density, then

$$\int_{\mathcal{X}} \phi(x)\, \rho(x)\; \mathrm{d}x \quad \text{should be read as} \quad \sum_{x \in X} \phi(x)\, \rho(x), \tag{1}$$

and represents the 'ρ-expected value' of ϕ, sometimes denoted $\mathbb{E}_\rho(\phi)$. However, if $\mathcal{X}$ is a measurable subset of $\mathbb{R}^N$, then this integral should be read as integration with respect to the Lebesgue measure (and in this case, ρ and F are always assumed to be Lebesgue-measurable functions).[3] For any $\mathcal{Y} \subseteq \mathcal{X}$, let $|\mathcal{Y}| := \int_{\mathcal{Y}} 1 \ \mathrm{d}y$. (So if $\mathcal{X}$ is finite or countable, then $|\mathcal{Y}|$ is just the cardinality of $\mathcal{Y}$.) A function $\pi : \mathcal{X} \longrightarrow \mathcal{X}$ is *measure-preserving* if π is almost-everywhere injective, and $|\pi^{-1}(\mathcal{Y})| = |\mathcal{Y}|$ for all $\mathcal{Y} \subseteq \mathcal{X}$. (If $\mathcal{X}$ is finite or countable, this just means that π is bijective —that is, π is a *permutation* of $\mathcal{X}$.) Let $\Pi_{\mathcal{X}}$ denote the set of all measure-preserving maps from $\mathcal{X}$ to itself. (So if $\mathcal{X}$ is finite, then $\Pi_{\mathcal{X}}$ is the corresponding permutation group.) Let $\Delta(\mathcal{X})$ denote the set of all probability density functions on the set $\mathcal{X}$.

1 Voting rules and estimators

Let $\mathcal{X}$ be a space of outcomes, and let $\mathcal{V}$ be a space of 'signals' or possible 'votes' which could be cast be each voter. If $\mathcal{I}$ is a set of voters, then a *profile* is an element $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$ which assigns a vote $v^i \in \mathcal{V}$ to each $i \in \mathcal{I}$. A *voting rule* is a multifunction $F : \mathcal{V}^{\mathcal{I}} \rightrightarrows \mathcal{X}$, which assigns to each profile $\mathbf{v}$ a nonempty subset $F(\mathbf{v}) \subseteq \mathcal{V}$.

For example, in the CJT, we have $\mathcal{X} = \mathcal{V} = \{\pm 1\}$, representing a decision about the truth/falsehood of a single statement. In *ordinal social choice*, there is a set $\mathcal{A}$ of 'social alternatives', and $\mathcal{X} = \mathcal{A}$, while $\mathcal{V}$ is the space of all strict preference orders over $\mathcal{A}$. Thus, each voter submits a preference order, and the rule selects one or more elements from $\mathcal{A}$. In *cardinal social choice*, $\mathcal{X} = \mathcal{A}$ and $\mathcal{V} \subseteq \mathbb{R}^{\mathcal{A}}$ is some set of possible cardinal utility functions on $\mathcal{A}$. In *Arrovian preference aggregation*, $\mathcal{X} = \mathcal{V}$ are both the space of preference orders over $\mathcal{A}$. In *judgement aggregation*, $\mathcal{V} = \mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$, where $\mathcal{K}$ is a set of statements, each of which could be either true or false, and $\mathcal{X}$ represents the set of logically possible truth-valuations of these statements.

Let $\Pi_{\mathcal{I}}$ be the group of all permutations of $\mathcal{I}$. For any $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$ and $\pi \in \Pi_{\mathcal{I}}$, we define $\pi(\mathbf{v}) := \mathbf{v}' \in \mathcal{V}^{\mathcal{I}}$ by setting $v'_i := v_{\pi(i)}$ for all $i \in \mathcal{I}$. A voting rule F is *anonymous* if $F(\pi(\mathbf{v})) = F(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$ and $\pi \in \Pi_{\mathcal{I}}$.

Estimators. Now, suppose $\mathcal{S}$ represents a space of possible states of nature, and let $s^* \in \mathcal{S}$ be the unknown true state. Again, let $\mathcal{V}$ be a space of possible 'messages' or 'signals' from voters. For all $i \in \mathcal{I}$, let $v^i \in \mathcal{V}$ be message indicating the beliefs of voter i about the true state; we regard v^i as a 'noisy signal' of s^*. We can then apply statistical techniques to find the best 'estimator' of s^* given the data $\{v^i\}_{i \in \mathcal{I}}$. For all $i \in \mathcal{I}$, $s \in \mathcal{S}$ and $v \in \mathcal{V}$, let $\rho^i_s(v)$ be the conditional probability that voter i will send the signal v, when the true state is in fact s; this defines a function $\rho : \mathcal{I} \times \mathcal{S} \longrightarrow \Delta(\mathcal{V})$, called the *error model*. The error model is *anonymous* if ρ does not depend on i —that is, $\rho^i_s(v) = \rho^j_s(v)$ for all $i, j \in \mathcal{I}$, $s \in \mathcal{S}$ and $v \in \mathcal{V}$. In this case, we can regard the error model as a function $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V})$.

Conditional on s^*, we assume the signals $\{v^i\}_{i \in \mathcal{I}}$ are independent random variables. Define the function $R : \mathcal{S} \longrightarrow \Delta(\mathcal{V}^{\mathcal{I}})$ by

$$R(s; \mathbf{v}) \quad := \quad \prod_{i \in \mathcal{I}} \rho^i_s(v^i), \qquad (2)$$

[3] All of our results can be generalized to arbitrary measures over arbitrary measure spaces, but it isn't really worth the extra technical overhead to work at that level of generality.

for any $s \in \mathcal{S}$ and $\mathbf{v} = (v^i)_{i\in\mathcal{I}} \in \mathcal{V}^{\mathcal{I}}$. Then $R(s;\mathbf{v})$ is the conditional probability of seeing the signal profile $\mathbf{v} := (v^i)_{i\in\mathcal{I}}$, given that the true state is s.

Let $\alpha \in \Delta(\mathcal{X})$ be an *a priori* probability density on $\mathcal{S}$, and let $\beta_{\mathbf{v}} \in \Delta(\mathcal{X})$ be the *a posteriori* distribution on $\mathcal{S}$, given the data $\mathbf{v}$. For all $s \in \mathcal{S}$, the value of $\beta_{\mathbf{v}}(s)$ can be computed using Bayes rule:

$$\beta_{\mathbf{v}}(s) \quad = \quad \frac{R(s;\mathbf{v})\,\alpha(s)}{\overline{R}(\mathbf{v})}, \quad \text{where} \quad \overline{R}(\mathbf{v}) \;:=\; \int_{\mathcal{S}} R(s;\mathbf{v})\,\alpha(s)\;\mathrm{d}s. \tag{3}$$

Finally, let $\mathcal{A}$ be a space of possible 'actions', and let $U : \mathcal{A}\times\mathcal{S}\longrightarrow\mathbb{R}$ be a utility function. Suppose our goal is to choose the action which will maximize expected utility, given our information about the unknown s^*. For any $a \in \mathcal{A}$, the *a posteriori* expected utility of action a, given $\mathbf{v}$ and α, is

$$\mathrm{EU}(a;\alpha,\mathbf{v}) \quad := \quad \int_{\mathcal{S}} U(a,s)\,\beta_{\mathbf{v}}(s)\;\mathrm{d}s \quad = \quad \frac{1}{\overline{R}(\mathbf{v})}\int_{\mathcal{S}} U(a,s)\,R(s;\mathbf{v})\,\alpha(s)\;\mathrm{d}s. \tag{4}$$

Thus, the *expected utility maximizer* (EUM) is the set

$$\mathrm{EUM}^{\mathcal{S},U}_{\alpha,\rho}(\mathbf{v}) \quad := \quad \operatorname*{argmax}_{a\in\mathcal{A}}\;\mathrm{EU}(a;\;\alpha,\mathbf{v}) \;\underset{(*)}{=}\; \operatorname*{argmax}_{a\in\mathcal{A}}\;\int_{\mathcal{S}} U(a,s)\,R(s;\mathbf{v})\,\alpha(s)\;\mathrm{d}s, \tag{5}$$

where $(*)$ is because the denominator $\overline{R}(\mathbf{v})$ in eqn.(4) is independent of the choice of $a \in \mathcal{A}$.

In some cases, we do not have a particular utility function in mind; we simply want to know the truth about s^*. The *maximum a posteriori* (MAP) *estimator* is the set of all $s \in \mathcal{S}$ which have maximal *a posteriori* probability:

$$\mathrm{MAP}^{\mathcal{S}}_{\alpha,\rho}(\mathbf{v}) \quad := \quad \operatorname*{argmax}_{s\in\mathcal{S}}\;\beta_{\mathbf{v}}(s) \;\underset{(*)}{=}\; \operatorname*{argmax}_{s\in\mathcal{S}}\;\Big(R(s;\mathbf{v})\,\alpha(s)\Big), \tag{6}$$

where $(*)$ is because the denominator $\overline{R}(\mathbf{v})$ in eqn.(3) is independent of s. (Equivalently, $\mathrm{MAP}^{\mathcal{S}}_{\alpha,\rho}(\mathbf{v}) = \mathrm{EUM}^{\mathcal{S},U}_{\alpha,\rho}(\mathbf{v})$, where we set $\mathcal{A} := \mathcal{S}$ and use the degenerate utility function U defined by $U(a,s) := 1$ if $s = a$ and $U(a,s) := 0$ if $s \neq a$.)

If we assume the prior distribution α is uniformly distributed over $\mathcal{S}$, then $\mathrm{MAP}^{\mathcal{S}}_{\alpha,\rho}(\mathbf{v})$ coincides with the *maximum likelihood estimator* (MLE), defined

$$\mathrm{MLE}^{\mathcal{S}}_{\rho}(\mathbf{v}) \quad := \quad \operatorname*{argmax}_{s\in\mathcal{S}}\;R(s;\mathbf{v}). \tag{7}$$

The CJT says that the majority voting rule is an MLE when $\mathcal{S} = \mathcal{V} = \{\pm 1\}$. The goal of this paper is to determine which other voting rules can function as EUM, MAP, or MLE for some plausible choice of U, ρ and α.

Scoring rules. Let $\overline{\mathbb{R}} := \mathbb{R}\cup\{-\infty\}$. Let $S : \mathcal{I}\times\mathcal{V}\times\mathcal{X}\longrightarrow\overline{\mathbb{R}}$ be a function. The *scoring rule* determined by S is the correspondence $F_S : \mathcal{V}^{\mathcal{I}} \rightrightarrows \mathcal{X}$ defined as follows. For all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$,

$$F_S(\mathbf{v}) \;:=\; \operatorname*{argmax}_{x\in\mathcal{X}}\;\overline{S}(\mathbf{v},x), \quad \text{where} \quad \overline{S}(\mathbf{v},x) \;:=\; \sum_{i\in\mathcal{I}} S^i(v^i,x), \;\text{for all } x \in \mathcal{X}. \tag{8}$$

In other words, for each $i \in \mathcal{I}$ and $x \in \mathcal{X}$, the vote v^i contributes $S^i(v^i, x)$ 'points' to the 'score' $\overline{S}(\mathbf{v}, x)$; we then choose the element(s) of $\mathcal{X}$ with the highest score. Note that, if $S^i(v, x) = -\infty$, then i voting for v is tantamount to i 'vetoing' alternative x; thus, we say F_S has *no vetos* if $S^i(v, x) > -\infty$ for all $(i, v, x) \in \mathcal{I} \times \mathcal{V} \times \mathcal{X}$. The scoring rule is *anonymous* if the score function S does not depend on $\mathcal{I}$ —that is, for all $i, j \in \mathcal{I}$, $v \in \mathcal{V}$, and $x \in \mathcal{X}$, we have $S^i(v, x) = S^j(v, x)$. In this case, we can treat S as a function $S : \mathcal{V} \times \mathcal{X} \longrightarrow \mathbb{R}$. A scoring rule F is *balanced* if $F = F_S$ for some scoring function S such that for all $i \in \mathcal{I}$ and all $x, y \in \mathcal{X}$, we have $\int_{\mathcal{V}} \exp[S^i(v, x)]\ \mathrm{d}v = \int_{\mathcal{V}} \exp[S^i(v, y)]\ \mathrm{d}v$. (This is a technical condition which can be seen as a weak form of 'neutrality'.)

If $\alpha \in \Delta(\mathcal{X})$ and $\rho : \mathcal{I} \times \mathcal{X} \longrightarrow \Delta(\mathcal{V})$, then the ordered pair (α, ρ) will be called a *scenario* on $\mathcal{I} \times \mathcal{X} \times \mathcal{V}$. A voting rule $F : \mathcal{V}^{\mathcal{I}} \rightrightarrows \mathcal{X}$ is *MAP-rationalizable* if there exists a scenario (α, ρ) such that $F(\mathbf{v}) = \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$. In particular, F is *MLE-rationalizable* if it is MAP-rationalizable with α being the uniform density on $\mathcal{X}$. An error model ρ has *no impossibilities* if $\rho_x(v) > 0$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$. Here is the first major result of the paper.

Theorem 1.1 *Let $F : \mathcal{V}^{\mathcal{I}} \rightrightarrows \mathcal{X}$ be a voting rule.*

- **(a)** *F is MAP-rationalizable if and only if F is a scoring rule.*
- **(b)** *F is MLE-rationalizable if and only if F is a balanced scoring rule.*

Now suppose F is a scoring rule.

- **(a)** *F is anonymous if and only if there exists an anonymous score function for F, if and only if there exists an anonymous error model which MAP-rationalizes F.*
- **(b)** *F has no vetos if and only if the error model of F has no impossibilities.*

Say that F is *anonymously MLE-rationalizable* if F is MLE-rationalizable for some anonymous error model. Thus, Theorem 1.1(a,b,c) together imply:

> *F is anonymously MLE-rationalizable if and only if F is an anonymous, balanced scoring rule.*

The proof of Theorem 1.1(a) "$\Longrightarrow$" is based on identifying the expression $\overline{S}(x, \mathbf{v})$ in eqn.(8) with the the logarithm of the expression $R(x; \mathbf{v})\, \alpha(x)$ in eqn.(6). The logarithm converts the product (2) into a sum, which we can reformulate as a sum of suitably defined scoring functions $\{S^i;\ i \in \mathcal{I}\}$. The proof of "$\Longleftarrow$" simply reverses this argument, by identifying $\exp[\overline{S}(x, \mathbf{v})]$ as the product $R(x; \mathbf{v})\, \alpha(x)$, which is then factored in terms of a suitable prior α and error model ρ. The details are in the Appendix. However, it will be useful to illustrate the argument for two familiar voting rules.

Example 1.2. Let $\mathcal{X}$ be a set of social alternatives. One of these alternatives is truly the 'best' alternative; call it x^*. The true identity of x^* is unknown; our goal is to discover it (thus, in this model, $\mathcal{S} = \mathcal{X}$).

(a) (*Borda rule*) Let $\mathcal{V} := \mathcal{P}_{\mathcal{RF}}(\mathcal{X})$ be the set of all strict preference orders over $\mathcal{X}$, and let $N := |\mathcal{X}|$. For any $x \in \mathcal{X}$ and $v \in \mathcal{V}$, let $S(v, x) = N - n$ if there are n alternatives ranked higher than x in the ordering v. (In particular, $S(v, x) = N - 1$ if x is the *best* alternative according to v.) Then F_S is the Borda rule.

To MLE-rationalize this rule, we suppose each voter is most likely to choose a preference order that judges x^* to be best, and less likely to choose a preference order where x^* has lower rank, with probability exponentially decreasing according to the rank of x^*. To be precise, let $\epsilon \in (0, 1)$, and for all $x \in \mathcal{X}$ and $v \in \mathcal{V}$, suppose that

$$\rho_x(v) \quad := \quad \frac{\epsilon^{S(v,x)}}{C}, \quad \text{where} \quad C \quad := \quad (N-1)!\frac{\epsilon^N - 1}{\epsilon - 1}. \tag{9}$$

This yields an anonymous error model $\rho : \mathcal{X} \longrightarrow \Delta(\mathcal{V})$; it is easy to verify that $F_S = \mathrm{MLE}_\rho^{\mathcal{X}}$. (See also Proposition 2.5.)

This MLE-rationalization of Borda is equivalent to the one given by Young (1986, p.117).[4] Young supposes that there is some $p \in \left(\frac{1}{2}, 1\right)$ such that, for all $i \in \mathcal{I}$, and all $y \in \mathcal{X} \setminus \{x^*\}$, voter i correctly recognizes that that $x^* \succ y$ with probability p, whereas she falsely believes $y \succ x^*$ with probability $1 - p$. Meanwhile, for any distinct $y, z \in \mathcal{X} \setminus \{x^*\}$, she has an equal probability of thinking $y \succ z$ or $z \succ y$ (and these events are all independent). If we set $\epsilon := \frac{1-p}{p}$, then error model (9) follows.

(b) (*Approval voting*) Let $\mathcal{V} := \{0, 1\}^{\mathcal{X}}$. (A typical element of $\mathcal{V}$ will be written as $\mathsf{v} = (v_x)_{x \in \mathcal{X}}$, where $v_x \in \{0, 1\}$ for all $x \in \mathcal{X}$.) For any $x \in \mathcal{X}$ and $\mathsf{v} \in \mathcal{V}$, if $v_x = 1$, then v 'approves' of alternative x, whereas if $v_x = 0$, then v 'does not approve' of x. The ***approval*** voting rule $\mathrm{Appr} : \mathcal{V}^{\mathcal{I}} \rightrightarrows \mathcal{X}$ chooses the alternative(s) which are 'approved' by the most voters (Brams and Fishburn, 1983). Formally, for any profile $\mathbf{v} := (\mathsf{v}^i)_{i \in \mathcal{I}}$ (where $\mathsf{v}^i \in \mathcal{V}$ for all $i \in \mathcal{I}$), we define $\mathrm{Appr}(\mathbf{v}) := \operatorname*{argmax}_{x \in \mathcal{X}} \sum_{i \in \mathcal{I}} v_x^i$.

Define $S(\mathsf{v}, x) := v_x$ for all $\mathsf{v} \in \mathcal{V}$ and $x \in \mathcal{X}$; then $\mathrm{Appr} = F_S$. We will show that Appr is anonymously MLE-rationalizable in terms of a very natural error model.

Let $p \in \left(\frac{1}{2}, 1\right]$. For each $i \in \mathcal{I}$ and $x \neq y \in \mathcal{X}$, we suppose

$$\begin{array}{c} \mathrm{Prob}\left[v_x^i = 1 \;\middle|\; x^* = x\right] \;=\; p \quad \text{and} \quad \mathrm{Prob}\left[v_x^i = 0 \;\middle|\; x^* = x\right] \;=\; 1 - p, \\ \text{while} \quad \mathrm{Prob}\left[v_y^i = 1 \;\middle|\; x^* = x\right] \;=\; \tfrac{1}{2} \;=\; \mathrm{Prob}\left[v_y^i = 0 \;\middle|\; x = x^*\right]. \end{array} \tag{10}$$

We further assume that the random variables $\{v_y^i;\ y \in \mathcal{X} \setminus \{x\},\ i \in \mathcal{I}\}$ are jointly independent, conditional on $x^* = x$. So, if x is the best alternative, then each voter has a better-than-50% chance of approving of x, while her approvals of the other alternatives are generated by independent fair coin flips.

Define $c := p/(1-p)$ (so $c > 1$), let $N := |\mathcal{X}|$, and define $M := 2^{N-1}(c + 1)$. It is easy to check that (10) corresponds to the anonymous error model $\rho : \mathcal{X} \longrightarrow \Delta(\mathcal{V})$ defined by

[4] Young (1988, p.1238 and 1997, §5) provides a different and more complicated MLE-rationalization of Borda. Young (1988) speculates that Condorcet (1785) probably understood —or at least, suspected —the MLE-rationalizability of the Borda rule, but he ignored it, so as to snub his rival.

$\rho_x(\mathsf{v}) := c^{v_x}/M$ for all $x \in \mathcal{X}$ and $\mathsf{v} \in \mathcal{V}$ (recall that $v_x \in \{0,1\}$, so $c^{v_x} \in \{1, c\}$). For any profile $\mathbf{v} := (\mathsf{v}^i)_{i\in\mathcal{I}} \in \mathcal{V}^{\mathcal{I}}$ and $x \in \mathcal{X}$, observe that

$$c^{\sum_{i\in\mathcal{I}} v_x^i} = \prod_{i\in\mathcal{I}} c^{v_x^i} = M^I \prod_{i\in\mathcal{I}} \rho_x(v^i). \qquad \text{Thus,}$$

$$\text{Appr}(\mathbf{v}) = \underset{x\in\mathcal{X}}{\text{argmax}} \sum_{i\in\mathcal{I}} v_x^i = \underset{x\in\mathcal{X}}{\text{argmax}}\ c^{\sum_{i\in\mathcal{I}} v_x^i} = \underset{x\in\mathcal{X}}{\text{argmax}} \prod_{i\in\mathcal{I}} \rho_x(v^i) \underset{(*)}{=} \text{MLE}_\rho^{\mathcal{S}}(\mathbf{v}),$$

as desired. Here, $(*)$ is by formulae (2) and (7). (See Example 4.2 for another statistical interpretation of approval voting.) $\diamondsuit$

Note that the scenario which MAP-rationalizes a voting rule F is not unique. Two score functions $S, \widetilde{S} : \mathcal{I}\times\mathcal{V}\times\mathcal{X}\longrightarrow\mathbb{R}$ are *equivalent* if we have $F_S(\mathbf{v}) = F_{\widetilde{S}}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$. For example, if there exists some $r > 0$ and $q : \mathcal{I}\times\mathcal{V}\longrightarrow\mathbb{R}$ such that $\widetilde{S}^i(v, x) = r\, S^i(v, x) + q^i(v)$ for all $(i, v, x) \in \mathcal{I}\times\mathcal{V}\times\mathcal{X}$, then clearly $\widetilde{S}$ is equivalent to S. Such a 'linear' relationship between $\widetilde{S}$ and S is sufficient, but not necessary for equivalence; $\widetilde{S}$ will still be equivalent to S if $\widetilde{S}^i = r\, S^i(v, x) + q^i(v) + \epsilon^i(v, x)$, where r and q are as before, and $\epsilon : \mathcal{I}\times\mathcal{V}\times\mathcal{X}\longrightarrow\mathbb{R}$ is some sufficiently small 'perturbation' term. The set of all score functions equivalent to S forms a convex cone in the vector space $\mathbb{R}^{\mathcal{I}\times\mathcal{V}\times\mathcal{X}}$. The next result characterizes the amount of freedom we have in picking a scenario which MAP-rationalizes a given voting rule F.

Proposition 1.3 *Let (α, ρ) and $(\widetilde{\alpha}, \widetilde{\rho})$ be two scenarios on $\mathcal{I}\times\mathcal{X}\times\mathcal{V}$. For all $(i, v, x) \in \mathcal{I}\times\mathcal{V}\times\mathcal{X}$, define $S^i(v, x) := \log\left(\rho_x^i(v)\, \alpha(x)^{1/I}\right)$ and $\widetilde{S}^i(v, x) := \log\left(\widetilde{\rho}_x^i(v)\, \widetilde{\alpha}(x)^{1/I}\right)$. Then $\text{MAP}_{\alpha,\rho}^{\mathcal{X}} = \text{MAP}_{\widetilde{\alpha},\widetilde{\rho}}^{\mathcal{X}}$ if and only if the score functions S and $\widetilde{S}$ are equivalent.*

Let $\mathcal{V}^* := \bigsqcup_{n=1}^{\infty} \mathcal{V}^n$. A *variable population voting rule* is a correspondence $F^* : \mathcal{V}^* \rightrightarrows \mathcal{X}$. For all $I \in \mathbb{N}$, let F^I be the restriction of F^* to a rule on $\mathcal{V}^I$. We say that F^* is *anonymous* if F^I is anonymous for all $I \in \mathbb{N}$.

For any $\mathbf{n} \in \mathbb{N}^{\mathcal{V}}$, let $\|\mathbf{n}\| := \sum_{v\in\mathcal{V}} n_v$. Define $\mathbb{N}_{\text{fin}}^{\mathcal{V}} := \{\mathbf{n} \in \mathbb{N}^{\mathcal{V}};\ \|\mathbf{n}\| < \infty\}$. For any profile $\mathbf{v} \in \mathcal{V}^*$, we can define a vector $\mathbf{n}(\mathbf{v}) \in \mathbb{N}_{\text{fin}}^{\mathcal{V}}$ by setting $n(\mathbf{v})_w := \#\{i \in \mathbb{N};\ v_i = w\}$, for each $w \in \mathcal{V}$. This yields a surjection $\mathbf{n} : \mathcal{V}^*\longrightarrow\mathbb{N}_{\text{fin}}^{\mathcal{V}}$. The rule $F^* : \mathcal{V}^{\mathcal{I}}\longrightarrow\mathcal{X}$ is anonymous if and only if there exists some correspondence $f : \mathbb{N}_{\text{fin}}^{\mathcal{V}} \rightrightarrows \mathcal{X}$ such that $F^*(\mathbf{v}) = f(\mathbf{n}(\mathbf{v}))$ for all $\mathbf{v} \in \mathcal{V}^*$.

For example, let $S : \mathcal{V}\times\mathcal{X}\longrightarrow\mathbb{R}$ be an anonymous scoring function. Then we define the (anonymous, variable-population) *scoring rule* $F_S^* : \mathcal{V}^* \rightrightarrows \mathcal{X}$ by

$$F_S^*(\mathbf{v}) := \underset{x\in\mathcal{X}}{\text{argmax}}\ \overline{S}(\mathbf{v}, x), \text{ where } \overline{S}(\mathbf{v}, x) := \sum_{i=1}^{I} S(v^i, x), \quad \forall\ I \in \mathbb{N},\ \mathbf{v} \in \mathcal{V}^I,\ x \in \mathcal{X}.$$

Equivalently, we could define $f_S : \mathbb{N}_{\text{fin}}^{\mathcal{V}} \rightrightarrows \mathcal{X}$ by

$$f_S(\mathbf{n}) := \underset{x\in\mathcal{X}}{\text{argmax}}\ \overline{S}(\mathbf{n}, x), \text{ where } \overline{S}(\mathbf{n}, x) := \sum_{v\in\mathcal{V}} n_v\, S(v, x), \text{ for all } \mathbf{n} \in \mathbb{N}_{\text{fin}}^{\mathcal{V}} \text{ and } x \in \mathcal{X}.$$

An anonymous, variable-population rule F^* is (anonymously) *MLE-rationalizable* if there exists an anonymous error model $\rho : \mathcal{V}\longrightarrow\Delta(\mathcal{X})$ such that, for any profile $\mathbf{v} \in \mathcal{V}^*$, we have

$F^*(\mathbf{v}) = \mathrm{MLE}_\rho^{\mathcal{X}}(\mathbf{v})$. The next result extends Theorem 1.1 to variable populations, but it also provides a much tighter characterization of the rationalizing error model than the one given by Proposition 1.3.

Theorem 1.4 *Let $F^* : \mathcal{V}^* \rightrightarrows \mathcal{X}$ be an anonymous, variable-population voting rule.*

(a) *F^* is MLE-rationalizable if and only if F^* is a balanced scoring rule.*

(b) *Suppose that, for all $x, y \in \mathcal{X}$, there exists some $\mathbf{v} \in \mathcal{V}^*$ such that $F(\mathbf{v}) = \{x, y\}$. Let $\rho, \widetilde{\rho} : \mathcal{X} \longrightarrow \Delta(\mathcal{V})$ be two anonymous error models, and suppose $F = \mathrm{MLE}_\rho^{\mathcal{X}}$. Then $F = \mathrm{MLE}_{\widetilde{\rho}}^{\mathcal{X}}$ if and only if there is a constant $r > 0$ and a function $\tau : \mathcal{V} \longrightarrow \mathbb{R}_+$ such that $\widetilde{\rho}_x(v) = \tau(v) \cdot \rho_x(v)^r$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$.*

As an illustration of (a), the MLE-rationalizations of the Borda and approval voting rules (Example 1.2) carry over verbatim to the variable-population context. As an illustration of (b), suppose that, for all $x, y \in \mathcal{X}$, there exists a measure-preserving function[5] $\pi_{xy} : \mathcal{V} \longrightarrow \mathcal{V}$ such that $\rho_y = \rho_x \circ \pi_{xy}$. (Thus, the error model 'looks the same' for every $x \in \mathcal{X}$. For example, this is true for distance-based error models on homogeneous spaces; see Corollary 2.2 below.) Then, for any $r > 0$, there is a constant $T_r > 0$ such that $\int_{\mathcal{V}} \rho_x(v)^r \, \mathrm{d}v = T_r$, for all $x \in \mathcal{X}$. (In particular, $T_1 = 1$, because ρ_x is a probability density.) Define $\tau_r(v) := T_r^{-1}$ for all $v \in \mathcal{V}$. Then define $\rho^r : \mathcal{X} \longrightarrow \Delta(\mathcal{V})$ by $\rho_x^r(v) := \tau_r(v) \cdot \rho_x(v)^r = \rho_x(v)^r / T_r$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$. Then for any $r > 0$, the function ρ^r is an error model which MLE-rationalizes F. Note that, as $r \to 0$, the density ρ_x^r becomes almost uniformly distributed over the support of ρ_x. In particular, if $\rho_x(v) > 0$ for all $(x, v) \in \mathcal{X} \times \mathcal{V}$, then ρ_x^r converges to the uniform density on $\mathcal{V}$ (so reach voter receives an extremely 'noisy' signal of the true state). On the other hand, as $r \to \infty$, note that ρ_x^r concentrates almost all its mass on $\operatorname{argmax}_{v \in \mathcal{V}} \rho_x(v)$ (so each voter receives a 'high fidelity' signal).

Reinforcement. Let $\Pi_{\mathcal{V}}$ be the group of permutations of $\mathcal{V}$, and let $\Pi_{\mathcal{X}}$ be the group of permutations of $\mathcal{X}$. For any $\pi \in \Pi_{\mathcal{V}}$, and any profile $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$, define $\pi(\mathbf{v}) := (\pi(v_i))_{i \in \mathcal{I}} \in \mathcal{V}^{\mathcal{I}}$. A variable-population rule F^* is *neutral* if, for any $\pi \in \Pi_{\mathcal{X}}$, there exists some $\widetilde{\pi} \in \Pi_{\mathcal{V}}$ such that $F[\widetilde{\pi}(\mathbf{v})] = \pi[F(\mathbf{v})]$ for all $\mathbf{v} \in \mathcal{V}^*$. This means the rule treats all the alternatives equally; for any $x, x' \in \mathcal{X}$, and any profile $\mathbf{v} \in \mathcal{V}^*$ such that $x \in F^*(\mathbf{v})$, there exists a permuted profile $\mathbf{v}' \in \mathcal{V}^*$ such that $x' \in F^*(\mathbf{v}')$.

If $I, J \in \mathbb{N}$, and $\mathbf{v} \in \mathcal{V}^I$ and $\mathbf{w} \in \mathcal{V}^J$ are two profiles, then let $\mathbf{vw}$ denote the element of $\mathcal{V}^{I+J}$ obtained by concatenating $\mathbf{v}$ and $\mathbf{w}$ in the obvious way. The rule F^* satisfies *reinforcement*[6] if, for all $\mathbf{v}, \mathbf{w} \in \mathcal{V}^*$, we have $F^*(\mathbf{vw}) = F^*(\mathbf{v}) \cap F^*(\mathbf{w})$ whenever this intersection is nonempty. Interpretation: if two disjoint subpopulations (represented by $\mathbf{v}$ and $\mathbf{w}$) each regard every element of some subset $\mathcal{X}' \subset \mathcal{X}$ as optimal (i.e. if $\mathcal{X}' = F^*(\mathbf{v}) \cap F^*(\mathbf{w})$), then the combined population (represented by $\mathbf{vw}$) should also regard the elements of $\mathcal{X}'$ —and only these elements —as optimal.

Observe that $\mathbf{n}(\mathbf{vw}) = \mathbf{n}(\mathbf{v}) + \mathbf{n}(\mathbf{w})$. Thus, if F^* is anonymous, then F^* satisfies reinforcement if and only if, for all $\mathbf{v}, \mathbf{w} \in \mathbb{N}_{\mathrm{fin}}^{\mathcal{V}}$, we have $f(\mathbf{n}+\mathbf{m}) = f(\mathbf{n}) \cap f(\mathbf{m})$ whenever

[5] Recall: if $\mathcal{V}$ is finite or countable, then this is just any permutation of $\mathcal{V}$.

[6] Sometimes this property is called *separability* or *consistency*.

this intersection is nonempty. For example, any variable-population scoring rule satisfies reinforcement.

An anonymous, variable-population rule F^* satisfies ***overwhelming majority***[7] if, for any $\mathbf{n}_1, \mathbf{n}_2 \in \mathbb{N}^{\mathcal{V}}_{\text{fin}}$, there exists some $M \in \mathbb{N}$ such that, for all $m > N$, we have $f(\mathbf{n}_1 + M\,\mathbf{n}_2) \subseteq f(\mathbf{n}_2)$. Interpretation: if one sub-population of voters (represented by $M\,\mathbf{n}_2$) is much larger than another sub-population ($\mathbf{n}_1$), then the choice of the combined population should be determined by the choice of the large sub-population —except that the small sub-population may act as a 'tie-breaker' in some cases. For example, any variable-population scoring rule satisfies overwhelming majority. By combining Theorem 1.4 with a result of Myerson (1995), we obtain the following.

Corollary 1.5 *Suppose $\mathcal{X}$ and $\mathcal{V}$ are finite. Let $F^* : \mathcal{V}^* \rightrightarrows \mathcal{X}$ be a neutral, anonymous, variable-population voting rule. Then F^* is MLE-rationalizable if and only if F^* satisfies reinforcement and overwhelming majority.*

For example, it is clear that the Borda and approval voting rules (Example 1.2) satisfy neutrality, reinforcement and overwhelming majority.

Example 1.6. Let $\mathcal{A}$ be a finite set of alternatives, and let $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$ be the set of all strict preference orders over $\mathcal{A}$. A (variable-population) ***preference aggregator*** is a correspondence $F : \mathcal{P}_{\mathcal{RF}}(\mathcal{A})^* \rightrightarrows \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$. (Thus, in this model, $\mathcal{V} := \mathcal{X} := \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$.) The rule F^* ***respects unanimity*** if $F^*(\mathbf{v}) = x$ whenever $\mathbf{v} \in \mathcal{P}_{\mathcal{RF}}(\mathcal{A})^*$ is a 'unanimous' profile such that $v^i = x$ for all $i \in \mathcal{I}$. (This is a very natural requirement, if we regard F^* as a statistical estimator of the 'true' preference order over $\mathcal{A}$.) A ***score-based preference aggregator*** is determined by a score function $S : \mathcal{P}_{\mathcal{RF}}(\mathcal{A}) \times \mathcal{P}_{\mathcal{RF}}(\mathcal{A}) \longrightarrow \mathbb{R}$. For example, the Kemeny (1959) rule is an anonymous, neutral, score-based preference aggregator, where, for any $v, x \in \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$, $S(v, x)$ is the number of pairwise comparisons where v and x agree. Thus, Theorems 1.1(b,c) and 1.4(a) imply that the Kemeny rule is anonymously MLE-rationalizable. This was first observed by Young (1986, 1988, 1995, 1997) (see Example 3.4 for details). Thus, Corollary 1.5 implies that the Kemeny rule satisfies reinforcement and overwhelming majority. Indeed, Young and Levenglick (1978) have shown that the Kemeny rule is the *only* neutral, anonymous, variable-population preference aggregator which respects unanimity, satisfies reinforcement, also also satisfies a condition they call 'local independence of irrelevant alternatives' (LIIA). Thus, Corollary 1.5 implies: *the Kemeny rule is the only neutral, anonymous, variable-population preference aggregator rule which is MLE-rationalizable, respects unanimity, and satisfies LIIA.* ♢

Remarks. (a) Conitzer et al. (2009; Theorem 1) have proved a special case of Theorem 1.1(a) for anonymous, neutral, preference aggregators. Earlier, the "$\Longleftarrow$" direction had been proved by Conitzer and Sandholm (2005; Theorem 1); they also (2005; Lemma 1) proved a special case of Corollary 1.5 "$\Longrightarrow$" for anonymous, neutral, preference aggregators. These papers investigated the MLE-rationalizability of many common ordinal social choice

[7] Sometimes this is propertyy called *continuity* or the *Archimedean property.*

rules and preference aggregators. More recently, Xia et al. (2010) have investigated this problem in the special case when the space of social alternatives is a Cartesian product.

(b) Note that Theorem 1.1(a) (the MAP-rationalizability of unbalanced scoring rules) does not generalize cleanly to the variable-population setting. The reason is that the *a priori* density α which is imputed from an unbalanced score function S depends upon the population size. Proposition 3.2(b) (below) illustrates this.

2 Metric voting rules

The results of §1 are rather abstract. The fact that there *exists* a scenario which MAP-rationalizes a particular voting rule F does not imply that this scenario is very plausible. We will now show how the class of 'metric' voting rules are MAP-rationalized by plausible 'metric' scenarios. Throughout this section, we will implicitly assume that $\mathcal{V} = \mathcal{S} = \mathcal{X}$. Let d be a metric[8] on $\mathcal{X}$, and let $E : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be a bounded function (usually nonincreasing). Define $\overline{E} : \mathcal{X} \longrightarrow \mathcal{X}$ by

$$\overline{E}(x) \quad := \quad \int_{\mathcal{X}} E\left(d(x,y)\right)\, \mathrm{d}y, \qquad \text{for all } x \in \mathcal{X}. \tag{11}$$

Say E is a ***generator*** if E decays quickly enough that $\overline{E}(x)$ is finite for all $x \in \mathcal{X}$.[9] The ***metric error model generated by*** E on $(\mathcal{X}, d)$ is the function $\rho^E : \mathcal{X} \longrightarrow \Delta(\mathcal{X})$ defined

$$\rho^E_x(v) \quad := \quad \frac{E[d(x,v)]}{\overline{E}(x)}, \qquad \text{for all } x, v \in \mathcal{X}.^{10} \tag{12}$$

Now, for all $i \in \mathcal{I}$, let $L^i : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be a function (usually nondecreasing). (Roughly speaking, the greater the value of L^i, the more 'weight' voter i will have.) The ***metric voting rule*** defined on $(\mathcal{X}, d)$ by the system $\mathbf{L} := (L^i)_{i\in\mathcal{I}}$ is the function $\mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}} : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathcal{X}$ defined

$$\mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}(\mathbf{v}) \quad := \quad \operatorname*{argmin}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} L^i\left(d(x,v^i)\right), \qquad \text{for all } \mathbf{v} \in \mathcal{V}^{\mathcal{I}}. \tag{13}$$

Theorem 2.1 *Let $(\mathcal{X}, d)$ be a metric space. For all $i \in \mathcal{I}$, let E_i be a generator, define $\overline{E}_i : \mathcal{X} \longrightarrow \mathbb{R}$ as in eqn.(11), and define $\rho^i := \rho^{E_i} : \mathcal{X} \longrightarrow \Delta(\mathcal{X})$ as in eqn.(12). Let $L_i(r) := -\ln[E_i(r)]$ for all $r \in \mathbb{R}_+$ and $i \in \mathcal{I}$. For all $x \in \mathcal{X}$, define*

$$\alpha(x) \quad := \quad \frac{1}{C}\prod_{i\in\mathcal{I}} \overline{E}_i(x), \quad \textit{where} \quad C \quad := \quad \int_{\mathcal{X}} \prod_{i\in\mathcal{I}} \overline{E}_i(x)\, \mathrm{d}x. \tag{14}$$

(a) *$\alpha \in \Delta(\mathcal{X})$, and $\mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}(\mathbf{v}) = \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{X}^{\mathcal{I}}$.*

[8] Note that d is not necessarily related to the Euclidean metric, even if $\mathcal{X}$ is a subset of $\mathbb{R}^N$.

[9] As E is bounded, this is always true if $(\mathcal{X}, d)$ is bounded and $|\mathcal{X}| < \infty$. In particular, it holds automatically if $\mathcal{X}$ is finite, or if $\mathcal{X}$ is bounded in both the Euclidean metric and the d-metric.

[10] We divide by $\overline{E}(x)$ to ensure that ρ^E_x is a probability distribution on $\mathcal{X}$.

(b) *If* $\overline{E}_i$ *is a constant function for all* $i \in \mathcal{I}$*, then* $\mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}(\mathbf{v}) = \mathrm{MLE}^{\mathcal{X}}_{\rho}(\mathbf{v})$ *for all* $\mathbf{v} \in \mathcal{X}^{\mathcal{I}}$.

In most of our examples, the metric voting rule (13) is *anonymous*: there is some $L : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that $L^i = L$ for all $i \in \mathcal{I}$. Clearly, in Theorem 2.1, this occurs if and only if the error model ρ is anonymous, which means there is some generator $E : \mathbb{R}_+ \longrightarrow \mathbb{R}$ such that $E_i = E$ for all $i \in \mathcal{I}$. In this case, $\alpha(x) = \overline{E}(x)^I / C$, where $C = \int_{\mathcal{X}} \overline{E}^I(x)\, \mathrm{d}x$.

An *isometry* of $(\mathcal{X}, d)$ is a measure-preserving function[11] $f : \mathcal{X} \longrightarrow \mathcal{X}$ such that, for all $x, y \in \mathcal{X}$, we have $d[f(x), f(y)] = d(x, y)$. Let $\mathrm{Isom}(\mathcal{X}, d)$ be the group of all isometries of $(\mathcal{X}, d)$. Say that $(\mathcal{X}, d)$ is *homogeneous* if $\mathrm{Isom}(\mathcal{X}, d)$ acts transitively on $\mathcal{X}$; that is, for all $x, y \in \mathcal{X}$, there exists $f \in \mathrm{Isom}(\mathcal{X}, d)$ such that $f(x) = y$.

Corollary 2.2 *If* $(\mathcal{X}, d)$ *is homogeneous, then* $\overline{E}_i$ *is constant on* $\mathcal{X}$ *for all* $i \in \mathcal{I}$*; thus,* $\mathrm{MLE}^{\mathcal{X}}_{\rho} = \mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}$.

Example 2.3. (*Cyclic parameter estimation*) Fix $N \in \mathbb{N}$, and let $\mathcal{X} = [0 \ldots N)$ with the metric $d(x, y) = \min\{|x-y|, N-|x-y|\}$. Thus, $\mathcal{X}$ represents N points arranged uniformly around a circle. A vote over $\mathcal{X}$ represents an attempt to estimate some parameter ranging over this circle (e.g. an angle, a cyclical time unit such as day of the week). For all $m \in [0 \ldots N)$, define $F_m : \mathcal{X} \longrightarrow \mathcal{X}$ by $F_m(n) := (n + m) \bmod N$. Then F_m is an isometry of $\mathcal{X}$, and $\mathcal{X}$ is clearly homogeneous under this group of isometries. Thus, Corollary 2.2 says that, for any metric error model on $\mathcal{X}$, the MLE will be the corresponding metric voting rule. ♢

Example 2.4. (*Weighted plurality vote; the trivial metric*) Suppose $\mathcal{X}$ is finite, and $d(x, y) = 1$ for all $x \neq y$. (This represents an 'abstract' decision problem, with no structure on the space of alternatives.) This space is clearly homogeneous (every permutation is an isometry). For all $i \in \mathcal{I}$, if ρ^i is a metric error model (12), then there is some $\epsilon_i \in (0, 1)$ (measuring the 'error rate' of voter i) such that $\rho^i_x(y) = \epsilon_i / \overline{E}_i$ for all $x \neq y \in \mathcal{X}$, while $\rho^i_x(x) = 1/\overline{E}_i$. Here, $\overline{E}_i := (1 + (|\mathcal{X}| - 1)\epsilon_i)^{-1}$. In the notation of Theorem 2.1, we have $E_i(r) = \epsilon_i^r$ and thus, $L_i(r) = \lambda_i\, r$, where $\lambda_i := -\ln(\epsilon_i) > 0$ for all $i \in \mathcal{I}$. Thus, Corollary 2.2 says

$$\mathrm{MLE}^{\mathcal{X}}_{\rho}(\mathbf{v}) \quad = \quad \mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}(\mathbf{v}) \underset{(13)}{=\!=} \operatorname*{argmin}_{x \in \mathcal{X}} \sum_{i \in \mathcal{I}:\ v_i \neq x} \lambda_i, \quad \text{for all } \mathbf{v} \in \mathcal{V}^{\mathcal{I}}. \tag{15}$$

Define $\boldsymbol{\lambda} := (\lambda_i)_{i \in \mathcal{I}}$. Then (15) is clearly equivalent to the $\boldsymbol{\lambda}$*-weighted plurality* voting rule:

$$\mathrm{Plurality}^{\mathcal{X}}_{\boldsymbol{\lambda}}(\mathbf{v}) \quad := \quad \operatorname*{argmax}_{x \in \mathcal{X}} \sum_{i \in \mathcal{I}:\ v_i = x} \lambda_i, \quad \text{for all } \mathbf{v} \in \mathcal{V}^{\mathcal{I}}. \tag{16}$$

This MLE-rationalization of the weighted plurality rule is similar to Ben-Yashar and Paroush (2001, §4, eqn.(12)). In particular, if $\epsilon_i = \epsilon_j$ for all $i, j \in \mathcal{I}$ (all voters are equally competent), then (16) is just standard (anonymous) plurality voting rule. If $|\mathcal{X}| = 2$, then we obtain the CJT. If $|\mathcal{X}| \geq 3$, then we obtain a special case of List and Goodin (2001). ♢

[11] Recall: if $\mathcal{X}$ is finite or countable, then this is just any permutation of $\mathcal{X}$.

The quasigaussian error model, the mean rule, and Borda. Let $\mathcal{X} \subseteq \mathbb{R}^N$. Suppose the state of nature is a vector in $\mathcal{X}$, and the voters make observations of this vector, corrupted by Gaussian random noise. We can approximate this process with a *quasigaussian* error model, defined as follows. Let d_E be the Euclidean metric on $\mathcal{X}$. Let $\sigma > 0$ be a 'standard deviation' and define $E_\sigma(r) := \exp(-r^2/2\sigma^2)$ for all $r \in \mathbb{R}_+$ in eqn.(12). Thus, for all $\mathsf{x}, \mathsf{y} \in \mathcal{X}$, we have $\rho^\sigma_{\mathsf{x}}(\mathsf{y}) = \exp(-d_E(\mathsf{x}, \mathsf{y})^2/2\sigma^2)/\overline{E}(\mathsf{x})$, where $\overline{E}$ is defined as in (11).

In the notation of Theorem 2.1, we have $L(r) = r^2/2\sigma$ for all $r \in \mathbb{R}_+$. Thus, $\mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}$ is the *mean* voting rule, defined

$$\mathrm{Mean}_{\mathcal{X}}(\mathbf{v}) \; := \; \operatorname*{argmin}_{\mathsf{x}\in\mathcal{X}} \sum_{i\in\mathcal{I}} d_E(\mathsf{x}, \mathsf{v}^i)^2 \; \underset{(*)}{=} \; \operatorname*{argmin}_{x\in\mathcal{X}} \; d_E(\mathsf{x}, \overline{\mathsf{v}}), \tag{17}$$

$$\text{where} \quad \overline{\mathsf{v}} \; := \; \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \mathsf{v}_i, \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}^{\mathcal{I}}.$$

(Here $(*)$ is by a classic result of Christiaan Huygens, which states that $\sum_{i\in\mathcal{I}} d_E(\mathsf{x}, \mathsf{v}^i)^2 = |\mathcal{I}| \cdot d_E(\mathsf{x}, \overline{\mathsf{v}})^2 + \sum_{i\in\mathcal{I}} d_E(\overline{\mathsf{v}}, \mathsf{v}^i)^2$ for any $\mathsf{x} \in \mathbb{R}^N$.) Equation (17) yields an interesting interpretation of 'score-based' preference aggregators like the Borda Rule. Let $A \in \mathbb{N}$, let $\mathcal{A}$ be some set of A 'alternatives', and let $\mathcal{R} := \{r_1 < r_2 < r_3 < \cdots < r_A\} \subset \mathbb{R}$ be a set of A 'ranks'. A bijective function $\mathsf{v} : \mathcal{A} \longrightarrow \mathcal{R}$ is called a *ranking* of $\mathcal{A}$. Let $\mathcal{R}_{\mathcal{A}}$ be the set of all rankings of $\mathcal{A}$ (regarded as a subset of $\mathbb{R}^{\mathcal{A}}$). The $\mathcal{R}$*-scoring rule* is the voting rule $F_{\mathcal{R}} : (\mathcal{R}_{\mathcal{A}})^{\mathcal{I}} \rightrightarrows \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$ defined as follows. For any profile $\mathbf{v} \in (\mathcal{R}_{\mathcal{A}})^{\mathcal{I}}$, let $\overline{\mathsf{v}} := \left(\sum_{i\in\mathcal{I}} \mathsf{v}^i\right)/|\mathcal{I}|$ be its arithmetic mean (an element of $\mathbb{R}^{\mathcal{A}}$). Then $F_{\mathcal{R}}(\mathbf{v})$ is the set of all strict orderings $(\succ)$ of $\mathcal{A}$ such that, for all $a, b \in \mathcal{A}$, we have $(\overline{\mathsf{v}}(a) > \overline{\mathsf{v}}(b)) \Longrightarrow (a \succ b)$.[12] For example: if the elements of $\mathcal{R}$ are evenly spaced (e.g. $\mathcal{R} = \{1, 2, 3, \ldots, A\}$), then $F_{\mathcal{R}}$ is the *Borda* rule.[13]

Proposition 2.5 *Let $\rho^\sigma : \mathcal{R}_{\mathcal{A}} \longrightarrow \Delta(\mathcal{R}_{\mathcal{A}})$ be a quasigaussian error model. Then $\mathrm{MLE}^{\mathcal{R}_{\mathcal{A}}}_{\rho^\sigma}(\mathbf{v}) = \mathrm{Mean}_{\mathcal{R}_{\mathcal{A}}}(\mathbf{v})$ for any profile $\mathbf{v} \in (\mathcal{R}_{\mathcal{A}})^{\mathcal{I}}$. Thus, $F_{\mathcal{R}}(\mathbf{v})$ is the set of strict orderings of $\mathcal{A}$ determined by $\mathrm{MLE}^{\mathcal{R}_{\mathcal{A}}}_{\rho^\sigma}(\mathbf{v})$.*

Remarks. (a) Scoring rules like the Borda rule are usually seen as preference aggregators. Each voter i declares a preference order $(\succ_i)$ in $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$, and we 'impute' a quasi-cardinal ranking from $\mathcal{R}_{\mathcal{A}}$ to $(\succ_i)$ only as a computational device; the final output is another element of $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$. However, if voters know that their preferences will be aggregated using the $\mathcal{R}$-scoring rule, then they understand that, in declaring a preference order in $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$, they are *effectively* declaring a ranking in $\mathcal{R}_{\mathcal{A}}$. In the model of Proposition 2.5, we make this awareness explicit. Each voter attempts to perceive the 'true' ranking of the alternatives, but she is subject to idiosyncratic errors which are (roughly) independent normal random variables. The scoring rule is then the MLE of the 'true' ranking (and hence, the 'true' ordering) of $\mathcal{A}$. The problem with Proposition 2.5 is that the ranking

[12] Generically, the coordinates of $\overline{\mathsf{v}}$ are all distinct (hence strictly ordered); in this case, then $F_{\mathcal{R}}(\mathbf{v})$ is the *unique* element of $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$ which represents this ordering.

[13] In this case, the quasigaussian error model in $\mathcal{R}_{\mathcal{A}}$ is the Mallows (1957) *θ-model:* for any $\mathsf{x}, \mathsf{y} \in \mathcal{R}_{\mathcal{A}}$, we have $\rho^\sigma_{\mathsf{x}}(\mathsf{y}) = A \cdot B^{\mathrm{Sp}(\mathsf{x},\mathsf{y})}$, where $A, B > 0$ are constants, and $\mathrm{Sp}(\mathsf{x}, \mathsf{y})$ is the Spearman (1904) rank correlation between x and y; see Kendall (1970, p.101-102). (Compare this to footnote 17.)

system $\mathcal{R}$ seems totally arbitrary. It would seem more natural to let $\mathcal{R} = \mathbb{R}$ or $[0,1]$, or at least, some high-density subset of these spaces. However, the proof of Proposition 2.5 breaks down if $|\mathcal{R}| > |\mathcal{A}|$ (we lose homogeneity).

(b) In the setting of judgement aggregation (where $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}} \subseteq \mathbb{R}^{\mathcal{K}}$ for some set $\mathcal{K}$), the mean rule (17) is sometimes called the *fusion procedure*; its suitability as a statistical estimator has been studied by Hartman et al. (2011). The fusion procedure can be MAP-rationalized by positing a quasigaussian error model on $\{\pm 1\}^{\mathcal{K}}$, but the plausibility of such a model is debatable.

3 The exponential error model and the median rule

For all $i \in \mathcal{I}$, let $\epsilon_i \in (0,1)$ be the 'error rate' of voter i, and define $E_i(r) := \epsilon_i^r$ for all $r \in \mathbb{R}_+$. Thus, in eqn.(12), we have $\rho^{E_i} = \rho^{\epsilon_i}$, where for all $x, y \in \mathcal{X}$, we define $\rho_x^{\epsilon_i}(y) := \epsilon_i^{d(x,y)}/\overline{E}_i(x)$, with $\overline{E}_i$ defined as in (11). This is the *exponential error model*. For all $i \in \mathcal{I}$, let $\lambda_i := -\ln(\epsilon_i)$; then $\lambda_i > 0$ (because $\epsilon < 1$) and in Theorem 2.1, we have $L_i(r) = \lambda_i\, r$ for all $r \in \mathbb{R}_+$. Thus, $\mathrm{Min}\Sigma_{d,\mathbf{L}}^{\mathcal{X}}$ is the $(d, \boldsymbol{\lambda})$-*median* voting rule, defined

$$\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v}) \quad := \quad \operatorname*{argmin}_{x \in \mathcal{X}} \sum_{i \in \mathcal{I}} \lambda_i\, d(x, v^i), \quad \text{for all } \mathbf{v} \in \mathcal{V}^{\mathcal{I}}. \tag{18}$$

That is, $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v})$ is the set of elements in $\mathcal{X}$ minimizing the $\boldsymbol{\lambda}$-weighted average distance to the beliefs of the voters. In the anonymous case, $\epsilon_i = \epsilon_j$ (and hence, $\lambda_i = \lambda_j$) for all $i, j \in \mathcal{I}$; then $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v})$ simply minimizes the unweighted sum $\sum_{i \in \mathcal{I}} d(x, v^i)$.

Example 3.1. Let $(\mathcal{X}, d)$ be the trivial metric space from Example 2.4. Then any distance-based error model is an exponential error model, and the corresponding weighted median rule (18) is the weighted plurality rule (16). ◇

3.1 The interval

For any $N \in \mathbb{N}$, let $\mathcal{X}_N := \{\frac{n}{N};\ n \in [-N \ldots N]\}$, with the standard Euclidean metric d_E as a subset of $[-1,1]$. A vote over $\mathcal{X}_N$ thus represents an attempt to estimate some numerical parameter ranging over a 'discretized' version of the interval $[-1,1]$. Let $\mathbf{v} = (v^i)_{i \in \mathcal{I}} \in \mathcal{X}_N^{\mathcal{I}}$ be a profile, and suppose without loss of generality that $\mathcal{I} = [1...I]$ and $v^1 \leq v^2 \leq \cdots \leq v^i$. Then $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v})$ corresponds to the ordinary 'median' of the set $\{v^i\}_{i \in \mathcal{I}}$. That is: if I is odd (i.e. $I = 2J+1$ for some J), then $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v}) = v^J$, whereas if I is even (i.e. $I = 2J$), then $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v}) = \mathcal{X}_N \cap \left[v^J, v^{J+1}\right]$. For all $i \in \mathcal{I}$, suppose that voter i has a 'single-peaked' preference relation on $\mathcal{X}_N$, with the peak occuring at her ideal point v^i. Then Black (1948) showed that $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v})$ is the set of *Condorcet winners* in $\mathcal{X}_N$ —that is, the set of alternatives which can beat or tie every other alternative in a pairwise majority vote. Thus, $\mathrm{Median}_{d,\boldsymbol{\lambda}}^{\mathcal{X}}(\mathbf{v})$ will be the outcome of any Condorcet consistent voting rule.[14]

[14] Balinski and Laraki (2007, 2011) have recently analyzed the median rule on $\mathcal{X}_N$ in great detail.

We will now analyze the anonymous MAP-rationalization of the median rule on $\mathcal{X}_N$ using an anonymous exponential error model. First we introduce a useful notation. For any set $\mathcal{X}$ and any function $f : \mathcal{X} \longrightarrow \mathbb{R}_+$ such that $C := \int_{\mathcal{X}} f(x)\,\mathrm{d}x$ is finite, we define the probability density $\langle f \rangle_{\mathcal{X}} \in \Delta(\mathcal{X})$ by setting $\langle f \rangle_{\mathcal{X}}(x) := f(x)/C$ for all $x \in \mathcal{X}$. For example, equation (14) could be rewritten: $\alpha(x) := \left\langle \prod_{i \in \mathcal{I}} \overline{E}_i(x) \right\rangle_{\mathcal{X}}$, for all $x \in \mathcal{X}$. Likewise, the standard normal probability density on $\mathbb{R}$ is given by $\rho(r) := \left\langle e^{-r^2} \right\rangle_{\mathbb{R}}$, for all $r \in \mathbb{R}$.

Now, assume there is some $\epsilon > 0$ such that $E_i(r) := \epsilon^r$ for all $r \in \mathbb{R}_+$ and all $i \in \mathcal{I}$. The metric space $(\mathcal{X}_N, d_E)$ is not homogeneous, so Corollary 2.2 does not apply. However, Theorem 2.1(a) still tells us that $\mathrm{Median}^{\mathcal{X}}_{d,\boldsymbol{\lambda}}$ is the MAP estimator for a certain *a priori* probability density $\alpha_{\epsilon,N,I} \in \Delta(\mathcal{X}_N)$, defined:

$$\alpha_{\epsilon,N,I}(x) \underset{(14)}{:=} \left\langle \overline{E}_{\epsilon,N}(x)^I \right\rangle_{\mathcal{X}_N} \quad \text{where} \quad \overline{E}_{\epsilon,N}(x) \underset{(11)}{:=} \sum_{y \in \mathcal{X}_N} \epsilon^{d(x,y)} \quad \text{for all } x \in \mathcal{X}_N. \tag{19}$$

(Note the dependency on ϵ, I and N). The next result says that, if I and N are large enough, then $\alpha_{\epsilon,N,I}$ looks like a normal distribution with mean 0 and very small variance. Thus, as a statistical estimator on $\mathcal{X}_N$, the median rule (and hence, *any* Condorcet-consistent voting rule) is heavily 'centre biased'. In particular, it is not even a crude approximation of the MLE for the exponential error model on an interval.

Proposition 3.2 *Fix* $\epsilon > 0$. *Let* $\sigma_\epsilon^2 := \dfrac{1-\epsilon}{\ln(\epsilon)^2 \cdot \epsilon}$.

(a) *Fix* $I \in \mathbb{N}$. *For all* $s \in [-1,1]$ *and* $N \in \mathbb{N}$, *let* $s_N \in \mathcal{X}_N$ *be the element of* $\mathcal{X}_N$ *closest to* s. *Then define*

$$\alpha_{\epsilon,I}(s) := \lim_{N \to \infty} 2N\, \alpha_{\epsilon,N,I}(s_N). \tag{20}$$

Then $\alpha_{\epsilon,I} : [-1,1] \longrightarrow \mathbb{R}_+$ *is a probability density on* $[-1,1]$.

(b) *For all* $s \in [-1,1]$, *we have* $\displaystyle \lim_{I \to \infty} \left\langle \alpha_{\epsilon,I}\left(\frac{s}{\sqrt{I}}\right) \right\rangle_{[-1,1]} = \left\langle \exp\left(\frac{-s^2}{2\sigma_\epsilon^2}\right) \right\rangle_{[-1,1]}$.

For example, if $\epsilon = 0.01$, then $\sigma_\epsilon^2 \approx 4.668139501$; thus, if, $N > 100$ and $I = 500$, then the *a priori* density $\alpha_{\epsilon,N,I}$ is a discrete approximation of a normal distribution[15] with mean 0 and variance 0.009336279. Virtually all the mass of this distribution is concentrated in a tiny interval around 0.

More generally, if $\mathcal{X}$ is a similar 'discrete' model of the D-dimensional cube $[-1,1]^D$ (with independent errors in different dimensions), then the limit distribution will be a D-dimensional normal distribution.

[15] Strictly speaking, a normal distribution is defined on $\mathbb{R}$, not on $[-1,1]$. But when the variance is this small, only a tiny fraction (i.e. 10^{-10000}) of the density's mass lies outside $[-1,1]$, so this is irrelevant for practical purposes.

3.2 Judgement aggregation

Let $\{\pm 1\}^{\mathcal{K}}$ be the *Hamming cube*; a typical element will be denoted $\mathsf{x} := (x_k)_{k\in\mathcal{K}}$, where $x_k \in \{\pm 1\}$ for all $k \in \mathcal{K}$. Let d_H be the *Hamming metric* on $\{\pm 1\}^{\mathcal{K}}$, defined: $d_H(\mathsf{x},\mathsf{v}) := \#\{k \in \mathcal{K};\ x_k \neq v_k\}$, for all $\mathsf{x},\mathsf{v} \in \{\pm 1\}^{\mathcal{K}}$. We intepret each element of $\mathcal{K}$ as representing some proposition which could be either true or false. An element of $\{\pm 1\}^{\mathcal{K}}$ represents a *judgement* on the truth or falsehood of each of these propositions. The *propositionwise majority* voting rule $\mathrm{Maj} : (\{\pm 1\}^{\mathcal{K}})^{\mathcal{I}} \rightrightarrows \{\pm 1\}^{\mathcal{K}}$ is defined as follows: for any profile $\mathbf{v} := (\mathsf{v}^i)_{i\in\mathcal{I}} \in (\{\pm 1\}^{\mathcal{K}})^{\mathcal{I}}$, and any $k \in \mathcal{K}$, define $m_k := \sum_{i\in\mathcal{I}} v_k^i$. Thus, $m_k \geq 0$ if and only if a majority of voters assert $v_k^i = 1$. Now define $\mathrm{Maj}(\mathbf{v}) := \{\mathsf{x} \in \{\pm 1\}^{\mathcal{K}};\ m_k \cdot x_k \geq 0\}$; this is the set of all judgements which agree with the majority on every proposition.

Let $\mathsf{x}^* := (x_k^*)_{k\in\mathcal{K}} \in \{\pm 1\}^{\mathcal{K}}$ be the (unknown) true judgement. The next result says that the exponential error model arises in the Hamming cube when voters make independent random errors on each coordinate of their judgement. (Part (a) is a straightforward computation. Part (b) can be derived from part (a) via Corollary 2.2, but it can be seen more directly by applying the classic CJT to each of the $\mathcal{K}$ dimensions independently.)

Proposition 3.3 **(a)** *Fix $\delta \in \left(0, \frac{1}{2}\right)$, let $\epsilon := \frac{\delta}{1-\delta}$, and let $\rho^\epsilon : \{\pm 1\}^{\mathcal{K}} \longrightarrow \Delta(\{\pm 1\}^{\mathcal{K}})$ be the exponential error model.*

For all $i \in \mathcal{I}$ and $k \in \mathcal{K}$, suppose $\mathrm{Prob}\,[v_k^i \neq x_k^*] = \delta$ *and* $\mathrm{Prob}\,[v_k^i = x_k^*] = (1-\delta)$, *and these events are independent for distinct i and k. Then* $\mathrm{Prob}\left[\mathsf{v}^i = \mathsf{v}\ \middle|\ \mathsf{x}^* = \mathsf{x}\right] = \rho_{\mathsf{x}}^\epsilon(\mathsf{v})$ *for all $\mathsf{x},\mathsf{v} \in \{\pm 1\}^{\mathcal{K}}$ and $i \in \mathcal{I}$.*

(b) $\mathrm{MLE}_{\rho^\epsilon}^{\{\pm 1\}^{\mathcal{K}}}(\mathbf{v}) = \mathrm{Maj}(\mathbf{v})$ *for all $\mathbf{v} \in (\{\pm 1\}^{\mathcal{K}})^{\mathcal{I}}$.*

A *judgement space* is a subset $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$. This arises when there are logical interdependencies between the propositions in $\mathcal{K}$, so that some judgements in $\{\pm 1\}^{\mathcal{K}}$ are logically impossible. Social choice over a judgement space is called *judgement aggregation* (List and Puppe, 2009). The propositionwise majority voting rule (Maj) often yields judgements outside of $\mathcal{X}$, making it unattractive as a judgement aggregation rule (List and Pettit, 2002). Fortunately, the median rule (18) is still well-behaved, when we endow $\mathcal{X}$ with the Hamming metric d_H (Miller and Osherson, 2009; Nehring et al., 2009).

If $\rho^\epsilon : \{\pm 1\}^{\mathcal{K}} \longrightarrow \Delta(\{\pm 1\}^{\mathcal{K}})$ is the exponential error model described in Proposition 3.3(a), then we can define a 'restricted' error model $\rho_{|\mathcal{X}}^\epsilon : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ as follows: for all $\mathsf{x},\mathsf{v} \in \mathcal{X}$, we have

$$\rho_{\mathsf{x}|\mathcal{X}}^\epsilon(\mathsf{v}) \ := \ \frac{\rho_{\mathsf{x}}^\epsilon(\mathsf{v})}{\overline{\rho}_{|\mathcal{X}}^\epsilon(\mathsf{x})}, \quad \text{where} \quad \overline{\rho}_{|\mathcal{X}}^\epsilon(\mathsf{x}) \ := \ \sum_{\mathsf{y}\in\mathcal{X}} \rho_{\mathsf{x}|\mathcal{X}}^\epsilon(\mathsf{y}). \tag{21}$$

In general, the error model (21) cannot be justified as in Proposition 3.3(b), in terms of voters making independent errors in each coordinate of $\mathcal{K}$. (A voter's errors *cannot* be independent, because $\mathcal{X}$ imposes logical relationships between different coordinates). Nevertheless, the exponential error model (21) is still a simple and plausible model of voter error. If $\mathcal{X}$ is homogeneous, then Corollary 2.2 says that $\mathrm{MLE}_{\rho_{|\mathcal{X}}^\epsilon}^{\mathcal{X}} = \mathrm{Median}_{d_H}^{\mathcal{X}}$.

Example 3.4. (*The Kemeny rule*) Let $\mathcal{A}$ be some set of alternatives, and let $\mathcal{K} := \mathcal{A}^2$. Then any $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$ represents a ***tournament*** (i.e. a complete, antisymmetric binary relation) ($\underset{\mathsf{x}}{\succ}$) on $\mathcal{A}$, where, for all $a, b \in \mathcal{A}$, we define $a \underset{\mathsf{x}}{\succ} b$ if and only if $x_{a.b} = 1$. Every tournament on $\mathcal{A}$ corresponds to a unique element of $\{\pm 1\}^{\mathcal{K}}$ in this way. Let $\mathcal{P}_{\mathcal{RF}}(\mathcal{A}) \subset \{\pm 1\}^{\mathcal{K}}$ be the set of all ***strict preference orderings*** (i.e. transitive tournaments) on $\mathcal{A}$. Classical Arrovian preference aggregation is simply judgement aggregation on $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$. In the Appendix, we show that $(\mathcal{P}_{\mathcal{RF}}(\mathcal{A}), d_H)$ is homogeneous; thus, Corollary 2.2 says that $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{P}_{\mathcal{RF}}(\mathcal{A})} = \mathrm{Median}_{d_H}^{\mathcal{P}_{\mathcal{RF}}(\mathcal{A})}$. But the d_H-median rule on $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$ is the Kemeny (1959) rule. This MLE-rationalization of the Kemeny rule was first discovered by Young (1986, 1988, 1995, 1997). [16,17] ♢

Example 3.5. (*Committee selection*) Suppose $\mathcal{K}$ represents a set of 'candidates'. Then any element of $\mathsf{x} \in \{\pm 1\}^{\mathcal{K}}$ represents the 'committee' $\{k \in \mathcal{K};\ x_k = 1\}$. A judgement space $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$ thus represents a set of possible committees satisfying certain constraints on size or membership. Judgement aggregation over $\mathcal{X}$ thus represents the problem of electing an admissible committee from the candidates in $\mathcal{K}$. For any $\mathsf{x} \in \{\pm 1\}^{\mathcal{K}}$, let $\|\mathsf{x}\| := \#\{k \in \mathcal{K}\ ;\ x_k = 1\}$ (the size of the committee represented by x).

(a) Fix $N \in [1...K)$, and let $\mathcal{C}_{\mathcal{OM}}(N) := \{\mathsf{x} \in \{\pm 1\}^{\mathcal{K}}\ ;\ \|\mathsf{x}\| = N\}$; that is, the set of all committees comprised of exactly N candidates. In the Appendix we show that $(\mathcal{C}_{\mathcal{OM}}(N), d_H)$ is homogeneous; thus, Corollary 2.2 says that $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{C}_{\mathcal{OM}}(N)} = \mathrm{Median}_{d_H}^{\mathcal{C}_{\mathcal{OM}}(N)}$.

(b) Let $\mathcal{C}_{\mathcal{OM}}(\mathrm{odd}) = \{\mathsf{x} \in \{\pm 1\}^{\mathcal{K}};\ \|\mathsf{x}\| \text{ is odd}\}$. Aggregation over $\mathcal{C}_{\mathcal{OM}}(\mathrm{odd})$ represents an attempt to elect a committee from $\mathcal{K}$ having any odd cardinality (presumably to avoid the possibility of ties when the committee votes). In the Appendix we show that $(\mathcal{C}_{\mathcal{OM}}(\mathrm{odd}), d_H)$ is homogeneous; thus, $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{C}_{\mathcal{OM}}(\mathrm{odd})} = \mathrm{Median}_{d_H}^{\mathcal{C}_{\mathcal{OM}}(\mathrm{odd})}$. ♢

Example 3.6. (*Partitions*) Let $M_1, M_2, \ldots, M_L \in \mathbb{N}$, and let $\mathcal{N}$ be a set with $|\mathcal{N}| = M_1 + M_2 + \cdots + M_L$. An $(M_1, \ldots, M_L)$-***partition*** on $\mathcal{N}$ is an equivalence relation $(\sim)$ which has exactly L equivalence classes $\mathcal{M}_1, \ldots, \mathcal{M}_L \subset \mathcal{N}$, such that $|\mathcal{M}_\ell| = M_\ell$ for all $\ell \in [1 \ldots L]$. (For example: if $M_1 = \cdots = M_L$, then this would be an unlabelled equipartition of $\mathcal{N}$.) Let $\mathcal{K} := \mathcal{N}^2$; then any binary relation $(\sim)$ on $\mathcal{N}$ (and in particular, any equivalence relation) corresponds to a unique element $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$ such that $x_{n,m} = 1$ if and only if $n \sim m$. Let $\mathcal{E}(M_1, \ldots, M_L) \subset \{\pm 1\}^{\mathcal{K}}$ be the space of all $(M_1, \ldots, M_L)$-partitions. In the Appendix, we show that $(\mathcal{E}(M_1, \ldots, M_L), d_H)$ is homogeneous; thus, Corollary 2.2 says that $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{E}(M_1,\ldots,M_L)} = \mathrm{Median}_{d_H}^{\mathcal{E}(M_1,\ldots,M_L)}$. ♢

In spite of Examples 3.4-3.6, most judgement spaces are *not* homogeneous. For example, unlike Example 3.5, most committee-selection spaces are not homogeneous. Also, unlike Example 3.6, the space of *all* equivalence relations on $\mathcal{N}$ is not homogeneous. Neither

[16] Young argues that Condorcet (1785) had discovered the Kemeny rule and its MLE interpretation, but was unable to clearly explain his ideas. See also Example 1.6.

[17] Note that, when restricted to $\mathcal{P}_{\mathcal{RF}}(\mathcal{A})$, the Hamming metric d_H is Kendall's (1938) metric, and the exponential error model is the Mallows (1957) ϕ-*model*. (Compare this to footnote 13.) In fact, the MLE for the Mallows ϕ-model was first derived by Feigin and Cohen (1978, §3.1), and had earlier been suggested by Hays (1960, p.332); however, unlike Young, these authors did not connect it with the Kemeny rule.

are most truth-functional spaces (which represent the logically consistent answers to a set of logically interconnected propositions),[18] convexity spaces (representing classification problems), or simplex spaces (which represent resource allocation problems). Thus, in general, the *a priori* density α which the median rule imputes upon a judgement space $\mathcal{X}$ (via Theorem 2.1) will *not* be the uniform density, so the median rule cannot be interpreted as MLE on $\mathcal{X}$. The density α will tend to give more mass to elements of $\mathcal{X}$ which are more 'central', meaning that they have the most other elements of $\mathcal{X}$ close to them in the Hamming metric. Furthermore, as in Proposition 3.2(b), the density α will become more and more 'concentrated' on these central elements as the number of voters becomes large. Thus, just on the interval, the median rule on a generic judgement space will be heavily biased towards these 'central' elements.

4 EUM-rationalizability

A voting rule $F : \mathcal{V}^{\mathcal{I}} \longrightarrow \mathcal{A}$ is *anonymously EUM-rationalizable* if there exists a set $\mathcal{S}$, a utility function $U : \mathcal{A} \times \mathcal{S} \longrightarrow \mathbb{R}$, an anonymous error model $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V})$, and an *a priori* probability density $\alpha \in \Delta(\mathcal{S})$, such that $F(\mathbf{v}) = \mathrm{EUM}^{\mathcal{S},U}_{\alpha,\rho}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$, as defined in eqn.(5).

Let $\mathcal{R} \subset \mathbb{R}$ be a finite or infinite set, and let $\mathcal{A}$ be a set of alternatives. Set $\mathcal{V} := \mathcal{R}^{\mathcal{A}}$ (a typical element denoted by $\mathsf{v} = (v_a)_{a \in \mathcal{A}}$, where $v_a \in \mathcal{R}$ for all $a \in \mathcal{A}$). We define the *$\mathcal{R}$-quasiutilitarian* voting rule $\mathrm{QU}_{\mathcal{R}} : \mathcal{V} \rightrightarrows \mathcal{A}$ as follows. for any profile $\mathbf{v} = (\mathsf{v}^i)_{i \in \mathcal{I}} \in (\mathcal{R}^{\mathcal{A}})^{\mathcal{I}}$,

$$\mathrm{QU}_{\mathcal{R}}(\mathbf{v}) \quad := \quad \operatorname*{argmax}_{a \in \mathcal{A}} \sum_{i \in \mathcal{I}} v^i_a \quad \subseteq \quad \mathcal{A}. \tag{22}$$

In other words, each vote $\mathsf{v} \in \mathcal{R}^{\mathcal{A}}$ is treated as an $\mathcal{R}$-valued 'cardinal utility function' on $\mathcal{A}$, and we choose the alternative(s) with the highest average utility. For example, the classic utilitarian voting rule is quasiutilitarian (with $\mathcal{R} = \mathbb{R}$). The *relative utilitarian* rule is quasiutilitarian, with $\mathcal{R} = [0, 1]$. Approval voting is quasiutilitarian ($\mathcal{R} = \{0, 1\}$). The Borda rule is *not* quasiutilitarian, because it imposes a constraint that the coordinates of the vote v^i must all take distinct values. However, suppose we set $\mathcal{R} = \{1, 2, \ldots, A\}$, and allow voters to assign the same rank to two or more alternatives (e.g. due to indifference); then we get a variant of the Borda rule (the *relaxed Borda rule*), which *is* quasiutilitarian.

We shall see that quasiutilitarian rules are EUM-rationalizable for a broad and plausible range of scenarios. In these scenarios, $\mathcal{S} := \mathbb{R}^{\mathcal{A}}$, and $U : \mathbb{R}^{\mathcal{A}} \times \mathcal{A} \longrightarrow \mathbb{R}$ is simply defined by $U(\mathsf{r}, a) := r_a$ for all $\mathsf{r} \in \mathbb{R}^{\mathcal{A}}$ and $a \in \mathcal{A}$. In other words, the unknown state of nature is just a vector $\mathsf{u}^* \in \mathbb{R}^{\mathcal{A}}$, encoding the 'true' utility of each alternative in $\mathcal{A}$. Let $\rho : \mathbb{R} \longrightarrow \Delta(\mathcal{R})$ be a one-dimensional error model (so $\rho_r \in \Delta(\mathcal{R})$ for all $r \in \mathbb{R}$), Define $\rho^{\mathcal{A}} : \mathbb{R}^{\mathcal{A}} \longrightarrow \Delta(\mathcal{R}^{\mathcal{A}})$ by setting $\rho^{\mathcal{A}}_{\mathsf{u}}(\mathsf{v}) \;:=\; \prod_{a \in \mathcal{A}} \rho_{u_a}(v_a)$ for all $\mathsf{u} \in \mathbb{R}^{\mathcal{A}}$ and $\mathsf{v} \in \mathcal{R}^{\mathcal{A}}$. Thus, we suppose that each voter's vote v^i is a noisy signal of the 'true' utility function u^*, but with independent, identically distributed ρ-random noise in each coordinate.

[18] The maximum likelihood approach to truth-functional judgement aggregation has been considered by List (2005), Fallis (2005), Bovens and Rabinowicz (2006) and Hartman et al. (2011).

Let $\alpha \in \Delta(\mathbb{R}^{\mathcal{A}})$ be an *a priori* density, which is symmetric under permutation of the coordinates. (Thus, all the alternatives in $\mathcal{A}$ are *a priori* interchangeable.) Let $\overline{\mathcal{R}} := \{r_1 + r_2 + \cdots + r_I;\ r_1, \ldots, r_I \in \mathcal{R}\}$, a subset of $\mathbb{R}$. Define $\bar{\mathsf{v}} := \sum_{i\in\mathcal{I}} \mathsf{v}^i$ (a vector in $\overline{\mathcal{R}}^{\mathcal{A}}$). For any $r \in \overline{\mathcal{R}}$, we define $E_{\alpha,\rho}(r) := \mathbb{E}\left[u^*_a \;\middle|\; \overline{v}_a = r\right]$. (This function does not depend on a, because we assume the same noise model for all coordinates.) This yields a function $E_{\alpha,\rho} : \overline{\mathcal{R}} \longrightarrow \mathbb{R}$. The scenario (α, ρ) is ***regular*** if the function $E_{\alpha,\rho} : \overline{\mathcal{R}} \longrightarrow \mathbb{R}$ is strictly increasing. The proof of the next result is straightforward.

Proposition 4.1 *If* (α, ρ) *is regular, then* $\mathrm{QU}_{\mathcal{R}} = \mathrm{EUM}^{\mathcal{S},U}_{\alpha,\rho}$.

Example 4.2. (*Approval voting*) Suppose $\mathcal{R} = \{0, 1\}$, so $\mathrm{QU}_{\mathcal{R}}$ is approval voting (Example 1.2(b)). This means the alternatives in $a \in \mathcal{A}$ can have one of two 'quality levels': either 'good' ($u^*_a = 1$) or 'bad' ($u^*_a = 0$). Our goal is then simply to select any 'good' alternative. (For example: $\mathcal{A}$ might be a set of candidates for some position; each candidate is either 'competent' or 'incompetent'. Or, $\mathcal{A}$ might be a set of possible solutions to some problem; each solution could either be 'successful' or 'unsuccessful'. The 'quality level' of a might not completely determine the utility which a will generate; however, perhaps the quality level is the only information we can obtain about a. Thus, u^*_a can also be interpreted as the *expected* utility of a, conditional on knowing whether a is 'good' or 'bad'.)

For each $i \in \mathcal{I}$ and $a \in \mathcal{A}$, suppose voter i receives a noisy signal v^i_a about the true quality u^*_a of alternative a, which is incorrect with probability $\delta \in \left(0, \frac{1}{2}\right)$. The errors are independent random variables. Thus, in this model, $\mathcal{V} = \mathcal{S} = \{0, 1\}^{\mathcal{A}}$, with a one-dimensional error model $\rho : \{0, 1\} \longrightarrow \Delta(\{0, 1\})$ given by $\rho_u(v) := \delta$ if $u \neq v$ and $\rho_u(v) := 1 - \delta$ if $u = v$. Finally, assume the true utilities $\{u^*_a\}_{a\in\mathcal{A}}$ are i.i.d. random variables; thus, there is some $p \in [0, 1]$ such that $\alpha[u^*_a = 1] = p$ for each $a \in \mathcal{A}$. Clearly, $\overline{\mathcal{R}} = \{0, 1, \ldots, N\}$, and for all $a \in \mathcal{A}$ and $n \in \overline{\mathcal{R}}$, we have $E_{\alpha,\rho}(n) := \mathrm{Prob}\left[u^*_a = 1 \;\middle|\; \sum_{i\in\mathcal{I}} v^i_a = n\right]$. In the Appendix, we show that (α, ρ) is regular; thus, Proposition 4.1 says that $\mathrm{Appr}(\mathbf{v}) = \mathrm{EUM}^{\mathcal{S},U}_{\alpha,\rho}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$. ◇

Example 4.3. (*Classic utilitarianism*) $\mathrm{QU}_{\mathbb{R}}$ is the classic utilitarian social choice function. For all $a \in \mathcal{A}$, suppose the true utility u^*_a is a normal random variable, and for each voter $i \in \mathcal{I}$, suppose the error $e^i_a := (v^i_a - u^*_a)$ is another, independent, normal random variable. Assume that $\{u^*_a\}_{a\in\mathcal{A}}$ are identically distributed, that $\{e^i_a\}_{a\in\mathcal{A}}$ are identically distributed, and that all these random variables are jointly independent. Formally, this means that α is a normal distribution on $\mathbb{R}$, and $\mathcal{R} := \mathbb{R}$, and there is another normal distribution ϵ on $\mathbb{R}$ (the density of the errors $\{e^i_a\}_{a\in\mathcal{A}}$), such that $\rho_u(v) = \epsilon(v - u)$ for all $(u, v) \in \mathbb{R} \times \mathbb{R}$. It is easy to show that the scenario (α, ρ) is regular (see Appendix); thus, $\mathrm{QU}_{\mathbb{R}}$ is the EUM for this error model. ◇

'Regularity' is a weak condition, which will be true for almost any scenario where the average vote $\bar{\mathsf{v}}$ is some kind of 'noisy signal' of u^*. Note that we do *not* require $\bar{\mathsf{v}}$ to be a particularly accurate or unbiased estimate of u^*. Indeed, the scenario (α, ρ) can be regular even if the random error $(u^*_a - \overline{v}_a)$ has a large bias and a large variance. Thus, $\mathrm{QU}_{\mathcal{R}}$ is EUM-rationalizable by a very broad range of noise models.

For example, the next result says: if we begin with a regular scenario, and transform the votes via a nondecreasing function, then the result will be another regular scenario. Formally: let $\mathcal{R}, \widetilde{\mathcal{R}} \subseteq \mathbb{R}$, and let $f : \mathcal{R} \longrightarrow \widetilde{\mathcal{R}}$ be a nondecreasing function. Let $\rho : \mathbb{R} \longrightarrow \Delta(\mathcal{R})$ be an error model. For all $x \in \mathbb{R}$, define $\widetilde{\rho}_x \in \Delta(\widetilde{\mathcal{R}})$ to be the distribution of $f(r)$, where r is a ρ_x-random variable on $\mathcal{R}$. Then $\widetilde{\rho} : \mathbb{R} \longrightarrow \Delta(\widetilde{\mathcal{R}})$ is another error model. We use the notation $f(\rho) := \widetilde{\rho}$.

Lemma 4.4 *Let $\mathcal{R}, \widetilde{\mathcal{R}} \subseteq \mathbb{R}$, and let (α, ρ) be a regular scenario on $\mathcal{R}$. If $f : \mathcal{R} \longrightarrow \widetilde{\mathcal{R}}$ is a nondecreasing function, and $\widetilde{\rho} = f(\rho)$, then the scenario $(\alpha, \widetilde{\rho})$ is also regular.*

Example 4.5. Suppose $\mathcal{R} = \mathbb{R}$, and let (α, ρ) be a regular scenario (e.g. Example 4.3).

(a) $\mathrm{QU}_{[0,1]}$ is the *relative utilitarian* social choice rule (also called *range voting*). So, let $\widetilde{\mathcal{R}} := [0, 1]$, and define $f : \mathbb{R} \longrightarrow [0, 1]$ by $f(r) := \max\{0, \min\{1, r\}\}$ for all $r \in \mathbb{R}$. Then f is nondecreasing. Thus, if $\widetilde{\rho} = f(\rho)$, then Lemma 4.4 says $(\alpha, \widetilde{\rho})$ is regular, so Proposition 4.1 says $\mathrm{QU}_{[0,1]}$ is an EUM for $(\alpha, \widetilde{\rho})$.

(b) Fix $N \in \mathbb{N}$, and let $\widetilde{\mathcal{R}} := [0 \ldots N]$. If $N = |\mathcal{A}| - 1$, then $\mathrm{QU}_{[0\ldots N]}$ is the relaxed Borda rule. If $N = 1$, then we get $\mathrm{QU}_{\{0,1\}}$ —i.e. Approval Voting. Let $\widetilde{\mathcal{R}} := [0 \ldots N]$, and define $f : \mathbb{R} \longrightarrow \widetilde{\mathcal{R}}$ by $f(r) := \max\{0, \min\{N, \lfloor r \rfloor\}\}$ for all $r \in \mathbb{R}$. Then f is nondecreasing. Thus, if $\widetilde{\rho} = f(\rho)$, then Lemma 4.4 says $(\alpha, \widetilde{\rho})$ is regular, so Proposition 4.1 says $\mathrm{QU}_{[0\ldots N]}$ is an EUM for $(\alpha, \widetilde{\rho})$. ♢

Conclusion

This paper shows that many common voting rules (e.g. scoring rules, distance-based rules) can be interpreted as a maximum *a priori* estimator or expected utility maximizer for a suitably chosen error model. If a particular error model provide a plausible description of the epistemic problem faced by the voters, then these results provide a strong argument for using the corresponding voting rule. However, this 'statistical rationalization' approach has several fundamental shortcomings.

First, this paper assumes that the mistakes made by different voters are independent random variables. This is highly unrealistic: voters often draw on common information sources and deliberate with one another. However, Conitzer and Sandholm (2005; Proposition 1) have shown that *any* voting rule can be MLE-rationalized, if we allow arbitrary correlations between voters.[19] Thus, to obtain a meaningful statistical interpretation of voting rules, we must impose some constraints on voter interdependencies. For example, Dietrich and Spiekerman (2011) have extended the CJT to a model where the information

[19]For example, Bühlmann and Huber (1963) begin with the model of preference aggregation from Example 3.4. They suppose that each $a \in \mathcal{A}$ has some 'true' quality level $\theta_a \in \mathbb{R}$, and that the voters' errors are correlated such a way that the probability of a majority asserting $a \succ b$ is exactly $(1 + e^{\theta_b - \theta_a})^{-1}$ —and these majorities are conditionally independent. Under this (implausible) error model, they show (Theorem 2) that the Copeland rule is the expected utility maximizer for a wide variety of utility functions. But they also show (Theorem 1) that this EUM-rationalizaton of Copeland holds *only* for implausible error models of this kind. Thus, it is unlikely to be useful in practice.

flow between voters is described using a causal network. Can this approach be applied to other epistemic social choice problems?

Second, and somewhat related, this paper completely neglects strategic considerations; it assumes that all voters will vote truthfully. But Austen-Smith and Banks (1996) demonstrated the possibility for strategic voting in the CJT, even when all voters have the same objectives, and differ only in their beliefs. (Hummel (2010) summarizes more recent literature on this phenomenon.) To what extent does strategic voting undermine the performance of a voting rule as a statistical estimator?

Third, the statistical rationalization approach begins with a familiar voting rule, and then contrives some probabilistic scenario to rationalize it as a statistical estimator *ex post facto*. But this is backwards. One should begin by specifying the scenario which best describes the epistemic problem faced by the voters, and then derive the correct statistical estimator for this scenario. This estimator may or may not turn out to be a familiar voting rule. For example, Remage and Thompson (1964, 1966) characterized the MLE preference aggregator under a general error model, and Kendall (1970) studied a variety of related problems. Cohen and Feigin (1978, §3.1) derived the MLE preference aggregator the Mallows (1957) ϕ-model in Example 3.4 (i.e. the Kemeny rule). This was extended by Fligner and Verducci (1986, 1988, 1990, 1993) and Lebanon and Lafferty (2002; §3.3), who developed MLEs for generalizations of the Mallows ϕ-model. More recently, Drissi-Bakhkhat (2002), Drissi-Bakhkhat and Truchon (2004), Truchon (2008), and Truchon and Gordon (2008, 2009) have analyzed the MAP and EUM preference aggregators under a logistic error model. But little work has been done on developing general-purpose statistical estimators in other social choice frameworks, such as judgement aggregation (except for footnote 18).

Finally, the statistical rationalization approach requires a fairly precise specification of the error model of the voters. But sometimes this is not possible. The Condorcet Jury Theorem and the EUM results of §4 are fairly robust to underspecification of the error model (e.g. Lemma 4.4), but the MAP-rationalization results are not (Proposition 1.3 and Theorem 1.4(b)). Furthermore, in some cases, it may simply be inappropriate to model the epistemic problem using probabilities. What is the best approach to epistemic social choice which does not rely on statistical estimation?

Appendix: Proofs

Proof of Theorem 1.1. (a) "$\Longrightarrow$" Suppose there exists an error model $\rho : \mathcal{I} \times \mathcal{X} \longrightarrow \Delta(\mathcal{V})$ and *a priori* probability $\alpha \in \Delta(\mathcal{X})$ such that $F(\mathbf{v}) = \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$. Let $I := |\mathcal{I}|$. For all $i \in \mathcal{I}$, $v \in \mathcal{V}$ and $x \in \mathcal{X}$, define $S^i(v,x) := \log\left(\rho^i_x(v) \cdot \alpha(x)^{1/I}\right)$. Then for any $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$,

$$F(\mathbf{v}) = \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v}) \underset{(*)}{=} \operatorname*{argmax}_{x\in\mathcal{X}} \alpha(x) \cdot \prod_{i\in\mathcal{I}} \rho^i_x(v^i) = \operatorname*{argmax}_{x\in\mathcal{X}} \prod_{i\in\mathcal{I}} \left(\alpha(x)^{1/I} \cdot \rho^i_x(v^i)\right)$$

$$= \operatorname*{argmax}_{x\in\mathcal{X}} \log\left(\prod_{i\in\mathcal{I}} \alpha(x)^{1/I} \cdot \rho^i_x(v^i)\right) = \operatorname*{argmax}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} \log\left(\alpha(x)^{1/I} \cdot \rho^i_x(v^i)\right)$$

$$= \operatorname*{argmax}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} S^i(v^i,x), \qquad \text{as desired. Here, } (*) \text{ is by equations (2) and (6).}$$

"$\Longleftarrow$" Suppose there exists a score function $S : \mathcal{I}\times\mathcal{V}\times\mathcal{X}\longrightarrow\overline{\mathbb{R}}$ such that $F = F_S$. For all $i\in\mathcal{I}$, $v\in\mathcal{V}$ and $x\in\mathcal{X}$, define $\widetilde{\rho}^i_x(v) := \exp\left(S^i(v,x)\right)$. Then, for $i\in\mathcal{I}$ and $x\in\mathcal{X}$, define $M^i(x) := \int_{\mathcal{V}} \widetilde{\rho}^i_x(v)\ \mathrm{d}v$. Finally, for all $i\in\mathcal{I}$, $v\in\mathcal{V}$ and $x\in\mathcal{X}$, define $\rho^i_x(v) := \widetilde{\rho}^i_x(v)/M^i(x)$. Thus, $\rho^i_x\in\Delta(\mathcal{V})$ for all $i\in\mathcal{I}$ and $x\in\mathcal{X}$.

Next, for all $x\in\mathcal{X}$, define $\widetilde{\alpha}(x) := \prod_{i\in\mathcal{I}} M^i(x)$. Let $M := \int_{\mathcal{X}} \widetilde{\alpha}(x)\ \mathrm{d}x$, and for all $x\in\mathcal{X}$, define $\alpha(x) := \widetilde{\alpha}(x)/M$. Thus, $\alpha\in\Delta(\mathcal{X})$. For all $\mathbf{v}\in\mathcal{V}^{\mathcal{I}}$ and $x\in\mathcal{X}$, observe that

$$\begin{aligned}
\exp\left(\sum_{i\in\mathcal{I}} S^i(v^i,x)\right) &= \prod_{i\in\mathcal{I}} \exp\left(S^i(v^i,x)\right) = \prod_{i\in\mathcal{I}} \widetilde{\rho}^i(v^i,x)\\
&= \prod_{i\in\mathcal{I}} \left(M^i(x)\cdot\rho^i(v^i,x)\right) = \widetilde{\alpha}(x)\cdot\prod_{i\in\mathcal{I}}\rho^i(v^i,x)\\
&= M\cdot\alpha(x)\cdot\prod_{i\in\mathcal{I}}\rho^i(v^i,x).
\end{aligned}$$

$$\begin{aligned}
\text{Thus,}\quad F(\mathbf{v}) &= \operatorname*{argmax}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} S^i(v^i,x) = \operatorname*{argmax}_{x\in\mathcal{X}} \exp\left(\sum_{i\in\mathcal{I}} S^i(v^i,x)\right)\\
&= \operatorname*{argmax}_{x\in\mathcal{X}} M\cdot\alpha(x)\cdot\prod_{i\in\mathcal{I}}\rho^i(v^i,x) = \operatorname*{argmax}_{x\in\mathcal{X}} \alpha(x)\cdot\prod_{i\in\mathcal{I}}\rho^i(v^i,x)\\
&\underset{(*)}{=} \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v}), \quad \text{as desired. Here, } (*) \text{ is by equations (2) and (6).}
\end{aligned}$$

(b) Adapt the proof of Theorem 1.4 below. (c) and (d) are straightforward. □

Proof of Proposition 1.3. This follows immediately from the proof of Theorem 1.1(a). □

Proof of Theorem 1.4. (a) "$\Longrightarrow$" Suppose there is an anonymous error model ρ such that, for every $\mathbf{v}\in\mathcal{V}^*$, we have $F^*(\mathbf{v}) = \mathrm{MLE}^{\mathcal{X}}_{\rho}(\mathbf{v})$. Define $S : \mathcal{V}\times\mathcal{X}\longrightarrow\overline{\mathbb{R}}$ by $S(v,x) := \log(\rho_x(v))$ for all $(x,v)\in\mathcal{X}\times\mathcal{V}$. Then by applying the natural logarithm to defining formulae (2) and (7), it is easy to see that

$$\mathrm{MLE}^{\mathcal{X}}_{\rho}(\mathbf{v}) = \operatorname*{argmax}_{x\in\mathcal{X}} \log\left(\prod_{i\in\mathcal{I}}\rho_x(v_i)\right) = \operatorname*{argmax}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} S(v_i,x) = F^*_S(\mathbf{v}),$$

for all $\mathbf{v}\in\mathcal{V}^*$. A straightforward computation shows that S is balanced.

"$\Longleftarrow$" Suppose there is a balanced, anonymous scoring function S such that, for every $\mathbf{v}\in\mathcal{V}^*$, we have $F^*(\mathbf{v}) = F^*_S(\mathbf{v})$. Define $\widetilde{\rho} : \mathcal{X}\times\mathcal{V}\longrightarrow\mathbb{R}_+$ by $\widetilde{\rho}_x(v) := \exp[S(v,x)]$. Since S is balanced, there is some constant $M > 0$ such that, for all $x\in\mathcal{X}$, we have

$\int_{\mathcal{V}} \widetilde{\rho}_x(v)\,\mathrm{d}v = M$. Thus, if we define $\rho_x(v) := \widetilde{\rho}_x(v)/M$ for every $x \in \mathcal{X}$, then $\rho_x \in \Delta(\mathcal{V})$. By applying the exponential map to defining formula (8), it is easy to see that

$$F^*_S(\mathbf{v}) \;=\; \operatorname*{argmax}_{x\in\mathcal{X}} \prod_{i\in\mathcal{I}} \widetilde{\rho}_x(v^i) \;=\; \operatorname*{argmax}_{x\in\mathcal{X}} M^I \prod_{i\in\mathcal{I}} \rho_x(v^i) \;=\; \mathrm{MLE}^{\mathcal{S}}_{\rho}(\mathbf{v}),$$

for all $\mathbf{v} \in \mathcal{V}^*$.

(b) Suppose $\rho, \widetilde{\rho} : \mathcal{X} \longrightarrow \Delta(\mathcal{V})$ are two anonymous error models which MLE-rationalize F^*. Let $S := \log \circ \rho$ and $\widetilde{S} := \log \circ \widetilde{\rho}$; then the proof of (a) shows that $F^* = F^*_S$ and $F^* = F^*_{\widetilde{S}}$. Thus, a result of Pivato (2011) implies that there is some $r > 0$ and some function $t : \mathcal{V} \longrightarrow \mathbb{R}$ such that $\widetilde{S}(v, x) = r\,S(v, x) + t(v)$ for all $v \in \mathcal{V}$ and $x \in \mathcal{X}$. But that implies that $\widetilde{\rho}_x(v) = \tau(v) \cdot \rho_x(v)^r$, where we define $\tau(v) := \exp(t(v)) > 0$ for all $v \in \mathcal{V}$. □

Proof of Corollary 1.5. "$\Longrightarrow$" If F^* is MLE-rationalizable, then Theorem 1.4(a) says F^* is a scoring rule. It is easy to verify that any scoring rule satisfies reinforcement and overwhelming majority.

"$\Longleftarrow$" If $\mathcal{X}$ and $\mathcal{V}$ are finite, and F^* is neutral and satisfies reinforcement and overwhelming majority, then a theorem of Myerson (1995) says that F^* is a scoring rule. Furthermore, if F^* is neutral, then it is easy to check that the score function must be balanced. Finally, F^* is anonymous by hypothesis. Thus, Theorem 1.4(a) implies that F^* is anonymously MLE-rationalizable. □

Proof of Theorem 2.1. (a) For all $x \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{X}^{\mathcal{I}}$, we have

$$R(x, \mathbf{v}) \underset{(\dagger)}{=} \prod_{i\in\mathcal{I}} \frac{E_i[d(x, v^i)]}{\overline{E}_i(x)} \underset{(\diamond)}{=} \frac{1}{C\,\alpha(x)} \prod_{i\in\mathcal{I}} \exp(-L_i[d(x, v^i)]), \qquad \text{so}$$

$$R(x; \mathbf{v}) \cdot \alpha(x) \;=\; C \prod_{i\in\mathcal{I}} \exp\left(-L_i[d(x, v^i)]\right) \;=\; C\, \exp\left(-\sum_{i\in\mathcal{I}} L_i[d(x, v^i)]\right), \quad (23)$$

$$\text{so} \quad \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v}) \underset{(6)}{=} \operatorname*{argmax}_{x\in\mathcal{X}} R(x; \mathbf{v}) \cdot \alpha(x) \underset{(23)}{=} \operatorname*{argmax}_{x\in\mathcal{X}} \exp\left(-\sum_{i\in\mathcal{I}} L_i[d(x, v^i)]\right)$$

$$\underset{(*)}{=} \operatorname*{argmin}_{x\in\mathcal{X}} \sum_{i\in\mathcal{I}} L_i[d(x, v^i)] \underset{(13)}{=} \mathrm{Min}\Sigma^{\mathcal{X}}_{d,\mathbf{L}}(\mathbf{v}).$$

Here, $(\dagger)$ comes from substituting eqn.(12) into eqn.(2). Next, $(\diamond)$ comes from eqn.(14), and the fact that $E_i(r) = \exp(-L_i(r))$ for all $r \in \mathbb{R}$. Finally, $(*)$ is because the exponential function is increasing.

(b) If $\overline{E}_i$ is constant for all $i \in \mathcal{I}$, then α is the uniform density, so $\mathrm{MLE}^{\mathcal{X}}_{\rho}(\mathbf{v}) = \mathrm{MAP}^{\mathcal{X}}_{\alpha,\rho}(\mathbf{v})$. □

Proof of Corollary 2.2. Fix $i \in \mathcal{I}$. Let $x, y \in \mathcal{X}$. If $f \in \mathrm{Isom}(\mathcal{X}, d)$ and $f(x) = y$, then it is easy to see that $\overline{E}_i(x) = \overline{E}_i(y)$. Thus, if $(\mathcal{X}, d)$ is homogeneous, then $\overline{E}_i$ must be constant, for all $i \in \mathcal{I}$. Now apply Theorem 2.1(b). □

Proof of Proposition 2.5. First we will show that the metric space $(\mathcal{R}_\mathcal{A}, d_E)$ is homogeneous. For any permutation $\pi : \mathcal{A} \longrightarrow \mathcal{A}$, we can define an isometry $\pi_* : \mathcal{R}_\mathcal{A} \longrightarrow \mathcal{R}_\mathcal{A}$ by $\pi_*(\mathsf{v})(a) := \mathsf{v}(\pi(a))$ for all $a \in \mathcal{A}$ and $\mathsf{v} \in \mathcal{R}_\mathcal{A}$. To see that $\mathcal{R}_\mathcal{A}$ is homogeneous, let $\mathsf{v}, \mathsf{w} \in \mathcal{R}_\mathcal{A}$, and define $\pi : \mathcal{A} \longrightarrow \mathcal{A}$ by $\pi(a) = \mathsf{v}^{-1} \circ \mathsf{w}(a)$ for all $a \in \mathcal{A}$. Then π is well-defined because v is bijective; and π is itself bijective because w is also bijective; hence π is a permutation of $\mathcal{A}$, so $\pi_* \in \mathrm{Isom}(\mathcal{X}, d)$. It is easy to verify that $\pi_*(\mathsf{v}) = \mathsf{w}$.

Thus, for any $\mathbf{v} \in \mathcal{R}_\mathcal{A}^\mathcal{I}$, Corollary 2.2 says $\mathrm{MLE}_\rho^\mathcal{S}(\mathbf{v}) = \mathrm{Min}\Sigma_{d,\mathbf{L}}^\mathcal{X}(\mathbf{v}) = \mathrm{Mean}_\mathcal{X}(\mathbf{v})$. □

Proof of Proposition 3.2. (a) Define $\overline{E}_\epsilon : [-1,1] \longrightarrow \mathbb{R}$ by $\overline{E}_\epsilon(s) := 2 - \epsilon^{1-s} - \epsilon^{1+s}$ for all $s \in [-1,1]$. Define $\overline{E}_{\epsilon,N}(s)$ as in eqn.(19).

Claim 1: *For any $N \in \mathbb{N}$ and $s \in [-1,1]$, define $\widetilde{E}_{\epsilon,N}(s) := \left(1 - \epsilon^{1/N}\right) \overline{E}_{\epsilon,N}(s_N)$. Then the sequence of functions $\{\widetilde{E}_{\epsilon,N}\}_{N=1}^\infty$ converges uniformly to the function $\overline{E}_\epsilon$ on $[-1,1]$, as $N \to \infty$.*

Proof: Let $s \in [-1,1]$, and suppose $s_N = n/N$ for some $n \in [-N \ldots N]$. Then

$$\begin{aligned}
\overline{E}_{\epsilon,N}(s_N) \underset{(19)}{=} \sum_{y \in \mathcal{X}_N} \epsilon^{d(s_N,y)} &= 1 + \sum_{r=1}^{N+|n|} \epsilon^{r/N} + \sum_{r=1}^{N-|n|} \epsilon^{r/N} \\
&= 1 + \epsilon^{1/N} \cdot \frac{1 - \epsilon^{(N+|n|)/N}}{1 - \epsilon^{1/N}} + \epsilon^{1/N} \cdot \frac{1 - \epsilon^{(N-|n|)/N}}{1 - \epsilon^{1/N}} \\
&= 1 + \frac{\epsilon^{1/N}}{1 - \epsilon^{1/N}} \cdot \left(2 - \epsilon^{1+|s_N|} - \epsilon^{1-|s_N|}\right), \quad \text{because } |s_N| = |n|/N.
\end{aligned}$$

$$\begin{aligned}
\text{Thus,} \quad \widetilde{E}_{\epsilon,N}(s_N) &= \left(1 - \epsilon^{1/N}\right) \cdot \overline{E}_{\epsilon,N}(s_N) \\
&= \left(1 - \epsilon^{1/N}\right) + \epsilon^{1/N} \cdot \left(2 - \epsilon^{1+|s_N|} - \epsilon^{1-|s_N|}\right) \\
&= 1 + \epsilon^{1/N} - \epsilon^{1+\frac{1}{N}+|s_N|} - \epsilon^{1+\frac{1}{N}-|s_N|}.
\end{aligned}$$

Meanwhile, $\overline{E}_\epsilon(s) = 2 - \epsilon^{1+s} - \epsilon^{1-s} = 2 - \epsilon^{1+|s|} - \epsilon^{1-|s|}$. Thus, for all $s \in [-1,1]$, we have

$$\begin{aligned}
\left|\overline{E}_\epsilon(s) - \widetilde{E}_{\epsilon,N}(s)\right| &= \left|2 - \epsilon^{1+|s|} - \epsilon^{1-|s|} - 1 - \epsilon^{1/N} + \epsilon^{1+\frac{1}{N}+|s_N|} + \epsilon^{1+\frac{1}{N}-|s_N|}\right| \\
&\underset{(\triangle)}{\leq} \left|1 - \epsilon^{1/N}\right| + \left|\epsilon^{1+\frac{1}{N}+|s_N|} - \epsilon^{1+|s|}\right| + \left|\epsilon^{1+\frac{1}{N}-|s_N|} - \epsilon^{1-|s|}\right| \\
&\leq \left|1 - \epsilon^{1/N}\right| + \epsilon^{1+|s|} \cdot \left|\epsilon^{\frac{1}{N}+|s_N|-|s|} - 1\right| + \epsilon^{1-|s|} \cdot \left|\epsilon^{|s|+\frac{1}{N}-|s_N|} - 1\right| \\
&\underset{(*)}{\leq} \left|1 - \epsilon^{1/N}\right| + \left|\epsilon^{\frac{1}{N}+|s_N|-|s|} - 1\right| + \left|\epsilon^{|s|+\frac{1}{N}-|s_N|} - 1\right|
\end{aligned}$$

$$\underset{(\dagger)}{\leq} \quad 2\lambda \cdot \frac{1}{N} + 2\lambda \cdot \left| \frac{1}{N} + |s_N| - |s| \right| + 2\lambda \cdot \left| |s| + \frac{1}{N} - |s_N| \right| \qquad \text{(with } \lambda := \ln(\epsilon))$$

$$\underset{(\triangle)}{\leq} \quad 2\lambda \cdot \left(\frac{1}{N} + \frac{1}{N} + \Big| |s_N| - |s| \Big| + \frac{1}{N} + \Big| |s| - |s_N| \Big| \right)$$

$$\underset{(\ddagger)}{=} \quad 2\lambda \cdot \left(\frac{1}{N} + \frac{1}{N} + \frac{1}{N} + \frac{1}{N} + \frac{1}{N} \right) \quad = \quad \frac{10\lambda}{N}.$$

Here, $(\triangle)$ is by the triangle inequality, while $(*)$ is because $\epsilon^{1+|s|} \leq 1$ and $\epsilon^{1-|s|} \leq 1$ because $0 \leq \epsilon \leq 1$ and because $-1 \leq s \leq 1$. To see $(\dagger)$, define $f(x) := \epsilon^x$; then $f(0) = 1$ and $f'(0) = \lambda$, so if x is close to 0, then $|1-\epsilon^x| = |f(0)-f(x)| \leq 2f'(0)\cdot|0-x| = 2\lambda|x|$. Finally, $(\ddagger)$ is because $\Big| |s_N| - |s| \Big| \leq \frac{1}{N}$.

This bound holds for all $s \in [-1,1]$. Thus, $\left\| \overline{E}_\epsilon - \widetilde{E}_{\epsilon,N} \right\|_\infty \leq 10\lambda/N \xrightarrow[N\to\infty]{} 0$, as desired. ◇ Claim 1

Claim 2: *For all $s \in [-1,1]$, we have $\alpha_{\epsilon,I}(s) = \left\langle \overline{E}_\epsilon(s)^I \right\rangle_{[-1,1]}$. (Thus, $\alpha_{\epsilon,I} \in \Delta[-1,1]$.)*

Proof: Fix $I \in \mathbb{N}$. We have

$$\frac{\left(1-\epsilon^{1/N}\right)^I}{2N} \int_{\mathcal{X}_N} \overline{E}_{\epsilon,N}(x)^I \, dx \underset{(\diamond)}{=} \frac{\left(1-\epsilon^{1/N}\right)^I}{2N} \sum_{n=-N}^{N} \overline{E}_{\epsilon,N}(n/N)^I \tag{24}$$

$$\underset{(*)}{=} \int_{-1}^{1} \widetilde{E}_{\epsilon,N}^I \, ds \xrightarrow[N\to\infty]{(\dagger)} \int_{-1}^{1} \overline{E}_\epsilon(s)^I \, ds. \tag{25}$$

Here, $(\diamond)$ is just the notational convention (1), $(*)$ is by the (piecewise constant) definition of $\widetilde{E}_{\epsilon,N}$, and $(\dagger)$ is because the sequence of functions $\{\widetilde{E}_{\epsilon,N}^I\}_{N=1}^\infty$ converges uniformly to the function $\overline{E}_\epsilon^I$ on $[-1,1]$ (by Claim 1). Thus, for any $s \in [-1,1]$,

$$2N\,\alpha_{N,\epsilon,I}(s_N) \underset{(19)}{=} \frac{\overline{E}_{\epsilon,N}(s_N)^I}{\frac{1}{2N}\int_{\mathcal{X}_N} \overline{E}_{\epsilon,N}(x)^I \, dx} = \frac{\left(1-\epsilon^{1/N}\right)^I \cdot \overline{E}_{\epsilon,N}(s_N)^I}{\frac{\left(1-\epsilon^{1/N}\right)^I}{2N}\int_{\mathcal{X}_N} \overline{E}_{\epsilon,N}(x)^I \, dx}$$

$$\xrightarrow[N\to\infty]{(*)} \frac{\overline{E}_\epsilon(s)^I}{\int_{-1}^{1} \overline{E}_\epsilon(s)^I \, ds} = \left\langle \overline{E}_\epsilon(s)^I \right\rangle_{[-1,1]}.$$

Thus, $\alpha_{\epsilon,I}(s) \underset{(20)}{:=} \lim_{N\to\infty} 2N\,\alpha_{\epsilon,N,I}(s_N) = \left\langle \overline{E}_\epsilon(s)^I \right\rangle_{[-1,1]}$,

as desired. (To see $(*)$, note that $\left(1-\epsilon^{1/N}\right)^I \cdot \overline{E}_{\epsilon,N}(s_N)^I = \widetilde{E}_{\epsilon,N}(s)^I \xrightarrow[N\to\infty]{} \overline{E}_\epsilon(s)^I$ by Claim 1, while the denominator converges by eqn.(25).) ◇ Claim 2

(b) We will adapt a standard proof of the Central Limit Theorem (Folland, 1984, Theorem 9.14, p.299). Let $\lambda := \ln(\epsilon)$. For all $s \in [-1,1]$, define

$$\psi_\epsilon(s) := \frac{\overline{E}_\epsilon(s)}{\overline{E}_\epsilon(0)} = \frac{2-\epsilon^{1-s}-\epsilon^{1+s}}{2-2\epsilon}. \tag{26}$$

$$\text{Then}\quad \psi'_\epsilon(s) \quad = \quad \frac{\lambda}{2-2\epsilon}\cdot\left(\epsilon^{1-s}-\epsilon^{1+s}\right),$$
$$\text{and}\quad \psi''_\epsilon(s) \quad = \quad \frac{-\lambda^2}{2-2\epsilon}\cdot\left(\epsilon^{1-s}+\epsilon^{1+s}\right).$$
$$\text{Thus,}\quad \psi_\epsilon(0) \;=\; 1,\quad \psi'_\epsilon(0) \;=\; 0,\quad\text{and}\quad \psi''_\epsilon(0) \;=\; \frac{-\lambda^2\cdot\epsilon}{1-\epsilon}.$$

Thus, Taylor's Theorem says that, for all $s \in [-1,1]$,

$$\psi_\epsilon(s) \quad = \quad 1-\frac{\lambda^2\cdot\epsilon}{2(1-\epsilon)}\cdot s^2 \;+\; o(s^2) \quad = \quad 1-\frac{s^2}{2\sigma_\epsilon^2} \;+\; o(s^2).$$

where $o(s^2)$ is some function of s such that $\lim_{s\to 0}\frac{o(s^2)}{s^2} \;=\; 0$. Thus, for all $s \in \left[-\sqrt{I},\sqrt{I}\right]$, we have

$$\psi_\epsilon\left(\frac{s}{\sqrt{I}}\right) \quad = \quad 1-\frac{s^2}{2I\sigma_\epsilon^2} \;+\; o\left(\frac{s^2}{I}\right).$$
$$\text{Thus,}\quad \ln\left[\psi_\epsilon\left(\frac{s}{\sqrt{I}}\right)\right] \quad = \quad \ln\left[1-\frac{s^2}{2I\sigma_\epsilon^2} \;+\; o\left(\frac{s^2}{I}\right)\right] \;\underset{(*)}{=}\; -\frac{s^2}{2I\sigma_\epsilon^2} \;+\; o\left(\frac{s^2}{I}\right).$$
$$\text{Thus,}\ \ln\left[\psi_\epsilon\left(\frac{s}{\sqrt{I}}\right)^I\right] \quad = \quad I\cdot\ln\left[\psi_\epsilon\left(\frac{s}{\sqrt{I}}\right)\right] \quad = \quad -\frac{s^2}{2\sigma_\epsilon^2} \;+\; I\cdot o\left(\frac{s^2}{I}\right)$$
$$\xrightarrow[I\to\infty]{(\dagger)} \quad -\frac{s^2}{2\sigma_\epsilon^2}.$$
$$\text{Thus,}\quad \psi_\epsilon\left(\frac{s}{\sqrt{I}}\right)^I \quad \xrightarrow[I\to\infty]{(\ddagger)} \quad \exp\left(-\frac{s^2}{2\sigma_\epsilon^2}\right).$$
$$\text{Thus,}\quad \overline{E}_\epsilon\left(\frac{s}{\sqrt{I}}\right)^I \quad \underset{(26)}{=} \quad (2-2\epsilon)^I\cdot\psi_\epsilon\left(\frac{s}{\sqrt{I}}\right)^I \quad \xrightarrow[I\to\infty]{} \quad (2-2\epsilon)^I\cdot\exp\left(-\frac{s^2}{2\sigma_\epsilon^2}\right).$$
$$\text{Thus,}\quad \left\langle \alpha_{\epsilon,I}\left(\frac{s}{\sqrt{I}}\right)\right\rangle_{[-1,1]} \quad \underset{(\diamond)}{=} \quad \left\langle \overline{E}_\epsilon\left(\frac{s}{\sqrt{I}}\right)^I\right\rangle_{[-1,1]} \quad \xrightarrow[I\to\infty]{} \quad \left\langle \exp\left(-\frac{s^2}{2\sigma_\epsilon^2}\right)\right\rangle_{[-1,1]},$$

as desired. Here, equality $(*)$ is because Taylor's Theorem says that $\ln(1+s) = s + o(s)$ for all $s \in [0,2]$. Limit $(\dagger)$ is by definition of $o(\bullet)$, and $(\ddagger)$ is because the exponential function is continuous on $\mathbb{R}$. Finally, $(\diamond)$ is by Claim 2. □

The Hamming cube $\{\pm1\}^{\mathcal{K}}$ admits two kinds of isometries:

- *Coordinate permutations.* Let $\pi : \mathcal{K}\longrightarrow\mathcal{K}$ be a permutation. Define the bijection $\pi_* : \{\pm1\}^{\mathcal{K}}\longrightarrow\{\pm1\}^{\mathcal{K}}$ by $\pi_*(\mathsf{x})_k := x_{\pi(k)}$ for all $\mathsf{x} = (x_k)_{k\in\mathcal{K}} \in \{\pm1\}^{\mathcal{K}}$ and $k \in \mathcal{K}$. Then π_* is an isometry.
- *Coordinate reflections.* For any $\mathsf{x},\mathsf{y} \in \{\pm1\}^{\mathcal{K}}$, define $\mathsf{x}\odot\mathsf{y} := \mathsf{z} \in \{\pm1\}^{\mathcal{K}}$ by $z_k := x_k\cdot y_k$, for all $k\in\mathcal{K}$. For any $\mathsf{x}\in\{\pm1\}^{\mathcal{K}}$, define the involution $F_{\mathsf{x}} : \{\pm1\}^{\mathcal{K}}\longrightarrow\{\pm1\}^{\mathcal{K}}$ by $F_{\mathsf{x}}(\mathsf{y}) := \mathsf{x}\odot\mathsf{y}$ for all $\mathsf{y}\in\{\pm1\}^{\mathcal{K}}$. Thus, F_{x} simply acts on $\{\pm1\}^{\mathcal{K}}$ by 'flipping' certain coordinates and leaving the rest alone; this map is also an isometry.

Let $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$, and let $f \in \mathrm{Isom}(\{\pm 1\}^{\mathcal{K}}, d_H)$. Then $f \in \mathrm{Isom}(\mathcal{X}, d_H)$ if and only if $f[\mathcal{X}] = \mathcal{X}$.

Proof of Example 3.4. For any permutation $\pi : \mathcal{A} \longrightarrow \mathcal{A}$, we can define a permutation $\widetilde{\pi} : \mathcal{K} \longrightarrow \mathcal{K}$ by $\widetilde{\pi}(a, b) = (\pi(a), \pi(b))$; and then define $\widetilde{\pi}_* : \{\pm 1\}^{\mathcal{K}} \longrightarrow \{\pm 1\}^{\mathcal{K}}$. To see that $\widetilde{\pi}_*(\mathcal{P}_{\mathcal{RF}}(\mathcal{A})) = \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$, suppose $\mathsf{x} \in \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$ represents the preference order "$a_1 \prec a_2 \prec \cdots \prec a_N$"; then $\widetilde{\pi}_*(\mathsf{x})$ represents the preference order "$\pi(a_1) \prec \pi(a_2) \prec \cdots \prec \pi(a_N)$". Thus, $\widetilde{\pi}_* \in \mathrm{Isom}(\mathcal{P}_{\mathcal{RF}}(\mathcal{A}), d_H)$ for all $\pi \in \Pi_{\mathcal{A}}$.

To see that $(\mathcal{P}_{\mathcal{RF}}(\mathcal{A}), d_H)$ is homogeneous, let $\mathsf{x}, \mathsf{y} \in \mathcal{P}_{\mathcal{RF}}(\mathcal{A})$; we need some $f \in \mathrm{Isom}(\mathcal{P}_{\mathcal{RF}}(\mathcal{A}), d_H)$ such that $f(\mathsf{x}) = \mathsf{y}$. Suppose x represents "$a_1 \prec a_2 \prec \cdots \prec a_N$" while y represents "$b_1 \prec b_2 \prec \cdots \prec b_N$' (where $N := |\mathcal{A}|$). Define permutation $\pi : \mathcal{A} \longrightarrow \mathcal{A}$ by $\pi(a_n) := b_n$ for all $n \in [1...N]$. We have already seen that $\widetilde{\pi}_* \in \mathrm{Isom}(\mathcal{P}_{\mathcal{RF}}(\mathcal{A}), d_H)$, and it is clear that $\widetilde{\pi}_*(\mathsf{x}) = \mathsf{y}$, as desired. □

Proof of Example 3.5. (a) If $\pi : \mathcal{K} \longrightarrow \mathcal{K}$ is any permutation, then it is clear that $\pi_*(\mathcal{C}_{\mathcal{OM}}(N)) = \mathcal{C}_{\mathcal{OM}}(N)$; thus, $\pi_* \in \mathrm{Isom}(\mathcal{C}_{\mathcal{OM}}(N))$. Furthermore, the group of coordinate permutations acts transitively on $\mathcal{C}_{\mathcal{OM}}(N)$. Thus, $\mathcal{C}_{\mathcal{OM}}(N)$ is homogeneous, so Corollary 2.2 says $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{COM}(N)} = \mathrm{Min}\Sigma_{d_H,L}^{\mathcal{COM}(N)} = \mathrm{Median}_{d_H}^{\mathcal{COM}(N)}$.

(b) For any $\mathsf{z} \in \{\pm 1\}^{\mathcal{K}}$, if $\|\mathsf{z}\|$ is even, then it is easy to see that the coordinate reflection F_{z} is an isometry of $\mathcal{C}_{\mathcal{OM}}(\mathrm{odd})$. For any $\mathsf{x}, \mathsf{v} \in \mathcal{C}_{\mathcal{OM}}(\mathrm{odd})$, let $\mathsf{z} := \mathsf{v} \odot \mathsf{x}$; then $\|\mathsf{z}\|$ is even and $F_{\mathsf{z}}(\mathsf{x}) = \mathsf{v}$. Thus, $\mathcal{C}_{\mathcal{OM}}(\mathrm{odd})$ is homogeneous, so Corollary 2.2 says $\mathrm{MLE}_{\rho^\epsilon}^{\mathcal{COM}(\mathrm{odd})} = \mathrm{Min}\Sigma_{d_H,L}^{\mathcal{COM}(\mathrm{odd})} = \mathrm{Median}_{d_H}^{\mathcal{COM}(\mathrm{odd})}$. □

Proof of Example 3.6. Let $\mathcal{E} := \mathcal{E}_{(M_1,\ldots,M_L)}$. For any permutation $\pi : \mathcal{N} \longrightarrow \mathcal{N}$, define $\widetilde{\pi}_* : \{\pm 1\}^{\mathcal{K}} \longrightarrow \{\pm 1\}^{\mathcal{K}}$ as in Example 3.4. If $\mathsf{x} \in \mathcal{E}$ represents the equivalence relation $(\sim)$, then $\widetilde{\pi}_*(\mathsf{x})$ represents the equivalence relation $(\approx)$ such that $(n \approx m) \iff (\pi(n) \sim \pi(m))$. Thus, $\widetilde{\pi}_* \in \mathrm{Isom}(\mathcal{E}, d_H)$ for all $\pi \in \Pi_{\mathcal{N}}$.

To see that $(\mathcal{E}, d_H)$ is homogeneous, let $\mathsf{x}, \mathsf{x}' \in \mathcal{E}$ represent equivalence relations $(\sim)$ and $(\sim')$. Let $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_L$ be the equivalence classes of $(\sim)$, and let $\mathcal{M}'_1, \mathcal{M}'_2, \ldots, \mathcal{M}'_L$ be the equivalence classes of $(\sim')$, where $|\mathcal{M}_\ell| = |\mathcal{M}'_\ell| = M_\ell$ for all $\ell \in [1 \ldots L]$. Let $\pi_\ell : \mathcal{M}_\ell \longrightarrow \mathcal{M}'_\ell$ be any bijection. Define $\pi := \pi_1 \sqcup \pi_2 \sqcup \cdots \sqcup \pi_L$. Then $\pi : \mathcal{N} \longrightarrow \mathcal{N}$ is well-defined (because $\mathcal{N} = \mathcal{M}_1 \sqcup \mathcal{M}_2 \sqcup \ldots \sqcup \mathcal{M}_L$) and bijective (because $\mathcal{N} = \mathcal{M}'_1 \sqcup \mathcal{M}'_2 \sqcup \ldots \sqcup \mathcal{M}'_L$), and clearly, $\widetilde{\pi}_*(\mathsf{x}) = \mathsf{x}'$. □

Proof of Example 4.2. For all $\mathbf{v} \in \mathcal{V}^{\mathcal{I}}$ and $a \in \mathcal{A}$, let $S_a(\mathbf{v}) := \sum_{i \in \mathcal{I}} v_a^i$. Then for all $n \in [0...N]$, we have $E_{\alpha,\rho}(n) := \mathrm{Prob}\left[u_a^* = 1 \;\middle|\; S_a(\mathbf{v}) = n\right]$. We will use Bayes theorem to get a formula for $E_{\alpha,\rho}(n)$.

If the value of u_k^* were known, then $\{v_k^i\}_{i\in\mathcal{I}}$ would be i.i.d random variables; thus, $S_a(\mathbf{v})$ would be a binomially distributed random variable. For all $n \in [0...I]$, we have

$$\begin{aligned}
\text{Prob}\left[S_a(\mathbf{v}) = n \,\middle|\, u_k^* = 1\right] &= \binom{I}{n}(1-\delta)^n \cdot \delta^{I-n} \\
\text{and}\ \ \text{Prob}\left[S_a(\mathbf{v}) = n \,\middle|\, u_k^* = 0\right] &= \binom{I}{n}(1-\delta)^{I-n} \cdot \delta^{n}. \\
\text{Thus,}\quad \text{Prob}\left[u_k^* = 1\ \&\ S_a(\mathbf{v}) = n\right] &= \text{Prob}\left[S_a(\mathbf{v}) = n \,\middle|\, u_k^* = 1\right] \cdot \text{Prob}\left[u_k^* = 1\right] \\
&= \binom{I}{n}(1-\delta)^n \cdot \delta^{I-n} \cdot p, \qquad (27)
\end{aligned}$$

and

$$\begin{aligned}
\text{Prob}\left[S_a(\mathbf{v}) = n\right] &= \text{Prob}\left[S_a(\mathbf{v}) = n\ \&\ u_k^* = 1\right] + \text{Prob}\left[S_a(\mathbf{v}) = n\ \&\ u_k^* = 0\right] \\
&= \binom{I}{n}(1-\delta)^n \cdot \delta^{I-n} \cdot p + \binom{I}{n}(1-\delta)^{I-n} \cdot \delta^n \cdot (1-p). \qquad (28)
\end{aligned}$$

Thus, Bayes Theorem says,

$$\begin{aligned}
E_{\alpha,\rho}(n) &= \frac{\text{Prob}\left[u_k^* = 1\ \&\ S_a(\mathbf{v}) = n\right]}{\text{Prob}\left[S_a(\mathbf{v}) = n\right]} \underset{(*)}{=} \frac{(1-\delta)^n \cdot \delta^{I-n} \cdot p}{(1-\delta)^n \cdot \delta^{I-n} \cdot p + (1-\delta)^{I-n} \cdot \delta^n \cdot (1-p)} \\
&= \frac{1}{1 + (1-\delta)^{I-2n} \cdot \delta^{2n-I} \cdot (1-p)/p} = \left(1 + \left(\frac{1-\delta}{\delta}\right)^{I-2n} \cdot \frac{1-p}{p}\right)^{-1}.
\end{aligned}$$

where $(*)$ is by equations (27) and (28). Now, $\left(\frac{1-\delta}{\delta}\right) > 1$ because $0 < \delta < \frac{1}{2}$. Thus, as n increases, the expression $\left(\frac{1-\delta}{\delta}\right)^{I-2n}$ decreases; hence $E_{\alpha,\rho}(n)^{-1}$ decreases; hence $E_{\alpha,\rho}(n)$ itself increases, as desired. □

Proof of Example 4.3. Let $I := |\mathcal{I}|$, let $\overline{e}_a := \frac{1}{I}\sum_{i\in\mathcal{I}} e_a^i$, and let $\overline{\mathsf{v}}' := \overline{\mathsf{v}}/I$. Then u_a^* and $\overline{e}_a$ are independent, normal random variables, so the random variables $\overline{v}'_a = u_a^* + \overline{e}_a$ and $\overline{w}_a := (u_a^* - \overline{e}_a)$ are also independent, normal random variables, and clearly, $u_a^* = (\overline{v}'_a + \overline{w}_a)/2$. Thus, for any $r \in \mathbb{R}$, if $r' := r/I$, then

$$E_{\alpha,\rho}(r) = \mathbb{E}\left[u_a^* \,\middle|\, \overline{v}'_a = r'\right] = \mathbb{E}\left[(r' + \overline{w}_a)/2 \,\middle|\, \overline{v}'_a = r'\right] \underset{(*)}{=} \frac{r' + \mathbb{E}\left[\overline{w}_a\right]}{2} = \frac{r/I + U - E}{2}$$

where $U := \mathbb{E}[u_a^*]$ (the mean of α) and $E := \mathbb{E}[\overline{e}_a]$ (the mean of ϵ). (Here, $(*)$ is because $\overline{v}_a$ and $\overline{w}_a$ are independent.) Thus, $E_{\alpha,\rho}(r)$ is an increasing (indeed, affine) function of r, so (α, ρ) is regular. □

Proof of Lemma 4.4. Define $\widehat{\alpha} \in \Delta(\mathcal{R})$ by $\widehat{\alpha}(r) := \int_{\mathbb{R}} \rho_x(r)\alpha(x)\,\mathrm{d}x$, for all $r \in \mathcal{R}$; this is the (unconditional) probability density of the noisy signal produced by the scenario (α, ρ). For all $\widetilde{r} \in \widetilde{\mathcal{R}}$, define $\underline{g}(\widetilde{r}) := \inf\{r \in \mathcal{R};\ \ f(r) = \widetilde{r}\}$ and $\overline{g}(\widetilde{r}) := \sup\{r \in \mathcal{R};\ f(r) = \widetilde{r}\}$. Then $\underline{g}(\widetilde{r}) \leq \overline{g}(\widetilde{r})$, and we have

$$\begin{aligned} E_{\alpha,\widetilde{\rho}}(\widetilde{r}) &= \frac{\displaystyle\int_{\underline{g}(\widetilde{r})}^{\overline{g}(\widetilde{r})} E_{\alpha,\rho}(r)\,\widehat{\alpha}(r)\,\mathrm{d}r}{\displaystyle\int_{\underline{g}(\widetilde{r})}^{\overline{g}(\widetilde{r})} \widehat{\alpha}(r)\,\mathrm{d}r} && \text{if } \underline{g}(\widetilde{r}) < \overline{g}(\widetilde{r}), \\ \text{and}\quad E_{\alpha,\widetilde{\rho}}(\widetilde{r}) &= E_{\alpha,\rho}(r) && \text{if } \underline{g}(\widetilde{r}) = \overline{g}(\widetilde{r}) = r \text{ for some } r \in \mathcal{R}. \end{aligned} \tag{29}$$

Now, for any $\widetilde{r}_1, \widetilde{r}_2 \in \mathcal{R}$, if $\widetilde{r}_1 < \widetilde{r}_2$, then $\overline{g}(\widetilde{r}_1) \leq \underline{g}(\widetilde{r}_2)$, and either $\underline{g}(\widetilde{r}_1) < \underline{g}(\widetilde{r}_2)$ or $\overline{g}(\widetilde{r}_1) < \overline{g}(\widetilde{r}_2)$. By hypothesis, the function $E_{\alpha,\rho}$ is strictly increasing. Thus, eqn.(29) implies that $E_{\alpha,\widetilde{\rho}}(\widetilde{r}_1) < E_{\alpha,\widetilde{\rho}}(\widetilde{r}_2)$. □

References

Austen-Smith, D., Banks, J., 1996. Information aggregation, rationality, and the Condorcet jury theorem. American Political Science Review, 34–45.

Balinski, M., Laraki, R., May 22 2007. A theory of measuring, electing, and ranking. Proceedings of the National Academy of Science 104 (21), 8720–8725.

Balinski, M., Laraki, R., 2011. Majority Judgment: Measuring, Ranking, and Electing. MIT Press, Boston, MA.

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

Black, D. S., 1948. On the rationale of group decision-making. J. Political Economy 56, 23–34.

Bohman, J., Rehg, W. (Eds.), 1997. Deliberative Democracy : Essays on Reason and Politics. MIT Press.

Bovens, L., Rabinowicz, W., 2006. Democratic answers to complex questions—an epistemic perspective. Synthese 150 (1), 131–153.

Brams, S. J., Fishburn, P. C., 1983. Approval voting. Birkhäuser Boston, Mass.

Bühlmann, H., Huber, P. J., 1963. Pairwise comparison and ranking in tournaments. Ann. Math. Statist. 34, 501–510.

Cohen, A., Feigin, P., 1978. On a model of concordance between judges. Journal of the Royal Statistical Society B 40 (2), 203–213.

Cohen, J., 1986. An epistemic conception of democracy. Ethics 97 (1), 26–38.

Condorcet, M. d., 1785. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris.

Conitzer, V., Rognlie, M., Xia, L., 2009. Preference functions that score rankings and maximum likelihood estimation. In: 21st International Joint Conference on Artificial Intelligence (IJCAI-09). Pasadena, CA,, pp. 109–115.

Conitzer, V., Sandholm, T., 2005. Common voting rules as maximum likelihood estimators. In: 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-05). Edinburgh, pp. 145–152.

Dietrich, F., Spiekerman, K., 2011. Epistemic democracy with defensible premises. (preprint).

Drissi-Bakhkhat, M., 2002. A statistical approach to the aggregation of votes. Ph.D. thesis, Université Laval, Québec.

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

Estlund, D., 1997. Beyond fairness and deliberation: The epistemic dimension of democratic authority. In: Bohman and Rehg (1997), pp. 173–204.

Fallis, D., June 2005. Epistemic value theory and judgment aggregation. Episteme 2 (1), 39–55.

Fligner, M. A., Verducci, J. S., 1986. Distance based ranking models. J. Roy. Statist. Soc. Ser. B 48 (3), 359–369.

Fligner, M. A., Verducci, J. S., 1988. Multistage ranking models. J. Amer. Statist. Assoc. 83 (403), 892–901.

Fligner, M. A., Verducci, J. S., 1990. Posterior probabilities for a consensus ordering. Psychometrika 55 (1), 53–63.

Fligner, M. A., Verducci, J. S. (Eds.), 1993. Probability models and statistical analysis for ranking data. Springer-Verlag.

Folland, G. B., 1984. Real Analysis. John Wiley & Sons, New York.

Hartman, S., Pigozzi, G., Sprenger, J., 2011. Relaible methods of judgement aggregation. (preprint).

Hays, W. L., 1960. A note on average tau as a measure of concordance. J. Amer. Statist. Assoc. 55, 331–341.

Hummel, P., 2010. Jury theorems with multiple alternatives. Soc. Choice Welf. 34 (1), 65–103.

Kaniovski, S., 2010. Aggregation of correlated votes and Condorcet's jury theorem. Theory and Decision 69 (3), 453–468.

Kemeny, J. G., Fall 1959. Math without numbers. Daedalus 88, 571–591.

Kendall, M., 1938. A new measure of rank correlation. Biometrika 30, 81–93.

Kendall, M. G., 1970. Rank Correlation Methods. Griffin, London.

Lebanon, G., Lafferty, J., 2002. Cranking: Combining rankings using conditional models on permutations. In: International Conference on Machine Learning. pp. 363–370.

List, C., June 2005. Group knowledge and group rationality: A judgment aggregation perspective. Episteme 2 (1), 25–38.

List, C., Goodin, R. E., 2001. Epistemic democracy: Generalizing the Condorcet Jury Theorem. Journal of Political Philosophy 9 (3), 277–306.

List, C., Pettit, P., 2002. Aggregating sets of judgements: an impossibility result. Economics and Philosophy 18, 89–110.

List, C., Puppe, C., 2009. Judgement aggregation: a survey. In: Oxford handbook of rational and social choice. Oxford University Press, Oxford, UK., pp. 457–482.

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

Miller, M. K., Osherson, D., 2009. Methods for distance-based judgment aggregation. Social Choice and Welfare 32 (4), 575–601.

Myerson, R. B., 1995. Axiomatic derivation of scoring rules without the ordering assumption. Soc. Choice Welf. 12 (1), 59–74.

Nehring, K., Pivato, M., Puppe, C., 2009. Condorcet-efficient judgement aggregation. (preprint).

Pivato, M., 2011. Variable-population scoring rules. (preprint).

Remage, Jr., R., Thompson, Jr., W. A., 1964. Rankings from paired comparisons. Ann. Math. Statist. 35, 739–747.

Remage, Jr., R., Thompson, Jr., W. A., 1966. Maximum-likelihood paired comparison rankings. Biometrika 53, 143–149.

Spearman, C., 1904. The proof and measurement of association between two things. American Journal of Psychology 15, 72–101.

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

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

Truchon, M., Gordon, S., 2009. Statistical comparison of aggregation rules for votes. Math. Social Sci. 57 (2), 199–212.

Xia, L., Conitzer, V., Lang, J., 2010. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In: 9th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-10). Toronto, pp. 399–406.

Young, H. P., 1986. Optimal ranking and choice from pairwise decisions. In: Grofman, B., Owen, G. (Eds.), Information pooling and group decision making. JAI press, Greenwich, CT, pp. 113–122.

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

Young, H. P., Winter 1995. Optimal voting rules. Journal of Economic Perspectives 9 (1), 51–64.

Young, H. P., 1997. Group choice and individual judgments. In: Mueller, D. C. (Ed.), Perspectives on public choice: A handbook. Cambridge UP, Cambridge, UK, Ch. 9, pp. 181–200.

Young, H. P., Levenglick, A., 1978. A consistent extension of Condorcet's election principle. SIAM J. Appl. Math. 35 (2), 285–300.