

Thema

UMR 8184

THéorie Économique, Modélisation et Applications

Thema Working Paper n° 2014-16
Université de Cergy Pontoise, France

"Asymptotic utilitarianism in scoring rules"

Marcus Pivato



September, 2014

Asymptotic utilitarianism in scoring rules

Marcus Pivato*

THEMA, Université de Cergy-Pontoise[†]
and Department of Mathematics, Trent University[‡]

September 16, 2014

Abstract

Given a large enough population of voters whose utility functions satisfy certain statistical regularities, we show that voting rules such as the Borda rule, approval voting, and evaluative voting have a very high probability of selecting the social alternative which maximizes the utilitarian social welfare function. We also characterize the speed with which this probability approaches one as the population grows.

Keywords: utilitarian; relative utilitarian; approval voting; Borda; scoring rule.

JEL class: D63; D71.

*marcuspivato@gmail.com

[†]33 Boulevard du Port, 95011 Cergy-Pontoise cedex, France

[‡]1600 West Bank Drive, Peterborough, Ontario, Canada

1 Introduction

As a theory of social justice, utilitarianism is quite appealing. But as a practical criterion for collective choice, it has serious shortcomings. To apply the utilitarian criterion, we not only need accurate information about the cardinal utility functions of all individuals in society; we must also know how to make cardinal interpersonal comparisons between these utility functions. These informational problems are compounded by two other factors. First, a person may strategically misrepresent her utility. Second, she may inadvertently misperceive her own utility function, either through lack of self-knowledge, or because she does not fully understand the long-term consequences of the policies under consideration. For these reasons (among others), collective decisions are almost never made by trying to explicitly ascertain the utility functions of the members of society (e.g. via a survey). Instead, collective decisions are usually made by voting.

A *scoring rule* is a particular kind of voting rule where each voter assigns a “score” to each alternative (with some constraints), and the alternative with the highest aggregate score wins. Well-known scoring rules include the Borda rule, the plurality and anti-plurality rules, evaluative (or “range”) voting, and approval voting. Since they involve maximizing a sum, scoring rules seem like a sort of “ersatz utilitarianism”. We will show that this is more than just a superficial formal resemblance. If the probability distribution of utility functions in a large society satisfies certain conditions, then we will show that a well-chosen scoring rule has a very high probability of selecting the alternative which maximizes the utilitarian social welfare function. For a sufficiently large population, this probability can be made arbitrarily close to certainty. Thus, with the right scoring rule, we can realize the utilitarian ideal, despite the informational problems described above.

The remainder of this paper is organized as follows. Section 2 considers evaluative voting. Section 3 considers approval voting. Section 4 considers rank scoring rules, such as the Borda rule or the plurality rule. Each section introduces one or more scenarios (described by hypotheses concerning the probability distribution of utility functions), and then, for each scenario, gives an asymptotic probability result. Appendix A reviews some results from Pivato (2014b) that are used in the other proofs. Appendix B contains the proofs of all the results in the paper.

Related literature. The results in this paper are complementary to those in Pivato (2014a,b). Like the present paper, Pivato (2014a) considers conditions under which ordinal voting rules maximize the utilitarian social welfare function (SWF) in a large population. But whereas this paper focuses on scoring rules, Pivato (2014a) focuses on Condorcet consistent rules such as the Copeland rule or an agenda of pairwise majority votes. Meanwhile, Pivato (2014b) considers a broader problem: how can we compute (and maximize) the utilitarian SWF when we have only very imprecise information about people’s utility functions and the correct system of interpersonal utility comparisons, and when people can be strategically dishonest? Under plausible conditions, Pivato (2014b) shows that, in a large population, we can accurately estimate the utilitarian SWF despite these difficulties. Indeed, this can be done in a strategy-proof way, using a modified version of the

Groves-Clarke pivotal mechanism.

We will evaluate ordinal voting rules from a utilitarian perspective. This approach was pioneered by Rae (1969), Taylor (1969) and Weber (1978). Rae and Taylor assumed that all voters had equal preference intensities over a dichotomous choice, given by independent, identically distributed (i.i.d.) random $\{0, 1\}$ -valued utility functions. In this setting, they showed that simple majority vote maximized the expected value of the utilitarian SWF, amongst all anonymous voting rules.¹ Weber (1978) considered a setting with many alternatives, and variable preference intensities. Assuming that the voters' utilities for the different alternatives were independent, uniformly distributed (i.u.d.) random variables, he sought the voting rule which maximized the expected value of the utilitarian SWF in a large population. He showed that the Borda rule was the optimal rule in the class of rank scoring rules.² The results in Section 4 can be seen as a major generalization of this early insight. In the case of exactly three alternatives, Weber (1978) also showed that the approval voting rule slightly outperforms the Borda rule; we will study the utilitarian efficiency of approval voting in Section 3.

Shortly after Weber's foundational work, Bordley (1983, 1985a) and Merrill (1984) used computer simulations to estimate the expected value of the utilitarian SWF for various voting rules. Bordley (1985b, 1986) computed utilitarian-optimal weighted majority voting schemes for dichotomous decisions with correlated voters. But there was no further utilitarian analysis of voting rules for the next twenty years.

Starting in 2005, a literature emerged on the utilitarian analysis of federal representative assemblies. Most of these papers focussed on dichotomous decisions, and assumed that the utility functions of the citizens were i.i.d. random variables. They asked: which voting rule will maximise the expected value of the utilitarian SWF?³ First, Beisbart et al. (2005) computationally compared the performance of seven benchmark rules, while Barberà and Jackson (2006) gave an exact formula for the utilitarian-optimal weighted majority rule in terms of the distribution of utility functions found within each state. Next, Beisbart and Bovens (2007) and Bovens and Hartmann (2007) investigated the consequences of different population-based weighting formulas with a mixture of theoretical analysis and computational results. Laruelle and Valenciano (2008, Ch.3; 2010, §7) provided a utilitarian rationale for the classic Penrose "square root" weighting formula. Macé and Treibich (2012) and Koriyama et al. (2013) derived analytical results in scenarios where voters have non-separable preferences over a series of dichotomies.⁴ In contrast to all the previously mentioned papers, Fleurbaey (2009) and Beisbart and Hartmann (2010) considered models

¹In fact, Rae and Taylor were interested in maximizing "responsiveness": the probability that the outcome agrees with the preference of a random individual. But if all voters have equal preference intensities, then maximizing "responsiveness" is equivalent to maximizing the utilitarian SWF. See also Badger (1972), Curtis (1972), Schofield (1972), Straffin (1977), and Dubey and Shapley (1979) for extensions of the Rae-Taylor theorem. Riley (1990) appears to have independently intuited some of the same conclusions. But he did not give any formal proofs.

²Weber called these "weighted ranking rules".

³Several of these papers also consider maximizing the expected value of the Rawlsian (maximin) SWF.

⁴This is captured in their model by supposing that the utility function of each state is a *concave* function of its frequency of victory in a long series of decisions.

with *correlated* voters. In Biesbart and Hartmann’s model, the profile of utility functions is drawn from a multivariate normal distribution, whereas Fleurbaey’s model is extremely general; the correlation structure between voters is completely arbitrary.⁵ Finally, Maaser and Napel (2014) used computer simulations to find utilitarian-optimal voting weights in a setting with three or more alternatives arranged on a line (with single-peaked preferences). The overall message of these papers is that, with i.i.d. voters, the expected value of utilitarian social welfare is generally maximized by a “degressive” weighted majority rule, where the weight of a nation is a sub-linear function (e.g. square root) of its population.⁶ With non-independent and/or non-identical voters, the optimal weights can depend on the correlations and preference intensities.

Aside from Maaser and Napel (2014), these papers only considered *dichotomies*.⁷ But other recent papers have considered three or more alternatives, in the context of direct democracy. Lehtinen (2007, 2008) used computer simulations to show that strategic voting often *improves* the utilitarian efficiency of the Borda rule and approval voting. Caragiannis and Procaccia (2011) estimated the “distortion” of the plurality, approval, and antiplurality voting rules—that is, the worst-case ratio between the utilitarian social welfare of the optimal alternative, and the utilitarian social welfare of the alternative which actually wins, where the worst case is computed over all possible profiles of “normalized” utility functions. (A utility function is “normalized” if it is positive and the utilities sum to one.) Procaccia and Caragiannis were particularly interested in the asymptotic growth rate of this distortion ratio as the number of voters and/or alternatives becomes large. They showed that, if voters randomly convert their cardinal utility functions into voting behaviours in a plausible way, then the expected distortion ratio grows surprisingly slowly. Their intended application was preference aggregation in a cooperating group of artificially intelligent agents (e.g. Mars rovers), but their results are also applicable to more traditional social welfare problems. This approach has recently been extended by Boutilier et al. (2012), who study the worst-case and average-case performance of randomized social choice rules.

Given a social welfare function W , and the probability distribution of the voters’ cardinal utility functions, Apesteguia et al. (2011) asked: what ordinal voting rule maximizes the expected value of W ? If W is the utilitarian SWF, and the voters’ utilities are i.i.d. random variables, they showed that the W -optimal rule is a *rank scoring rule* of the kind we consider in Section 4.⁸ In particular, if the voters’ utilities are i.i.d. random variables, then the W -optimal rule is the Borda rule. The results in Section 4 of this paper can be seen as complementary to those of Apesteguia et al. (2011); while they show that a certain scoring rule is better, on average, than any alternative voting rule, we show that, in a large

⁵Fleurbaey (2009) also provides a utilitarian analysis of direct democracy, including a far-reaching generalization of the Rae-Taylor theorem to a setting where voters may have different preference intensities and arbitrary correlations.

⁶There is also a vast literature on *voting power* in representative assemblies, which reaches similar conclusions; see e.g. Laruelle and Valenciano (2008) for a survey.

⁷Fleurbaey (2009;§7) notes that his approach can be applied to any number of issues, because it assigns a distinct weight to each voter for each binary comparison (proportional to her preference intensity) and thereby evades the Condorcet paradox. However, his weighting formula is very informationally demanding.

⁸Apesteguia et al. (2011) also consider the case when W is the maximin or maximax SWF.

population, it approaches *perfect* agreement with the utilitarian social choice.

Giles and Postl (2012) conducted a similar investigation for (A, B) -*voting rules*, a two-parameter family of rules introduced by Myerson (2002), which includes approval vote as well as rank scoring rules. Giles and Postl suppose there are three alternatives, whose utilities for each voter are privately known i.i.d. random variables on the interval $[0, 1]$. Unlike Apesteguia et al., they focus only on the utilitarian SWF, and they allow strategic voting. Giles and Postl first characterize the symmetric Bayesian Nash equilibrium (BNE) for the N -player strategic voting game for any $N \geq 2$. Then they numerically compute the expected value of the utilitarian SWF at the three-player BNE for various $(A, B) \in [0, 1]^2$ (where the three players' utilities are i.i.d. random variables drawn from either a uniform distribution or a beta distribution on $[0, 1]$). The results on approval voting in Section 3 of this paper can be seen as complementary to the findings of Giles and Postl (2012), but extended to an arbitrary number of alternatives and a large number of voters.

Kim (2014) pushes this investigation further. In a setting with three or more alternatives, and voters with independent (but not identically distributed) random utilities, he characterizes the rules which are *ex ante* Pareto efficient in the class of ordinal voting rules: they are “non-anonymous” rank scoring rules (where each voter has perhaps a different score vector). He further shows that, in a “neutral” environment (i.e. all alternatives are *ex ante* interchangeable), such rules are incentive compatible (i.e. truth-revealing in BNE). Special cases of Kim’s analysis are the rank scoring rules which maximize *ex ante* utilitarian social welfare over all ordinal rules. In particular, Kim observes that the rank scoring rules of Apesteguia et al. (2011) are incentive-compatible in the i.i.d. environment of their paper.⁹ He then constructs incentive-compatible voting rules which, in terms of the utilitarian SWF, are superior to any ordinal rule (in particular, any scoring rule), but which utilize only a limited amount of cardinal utility information from the voters.

Azrieli and Kim (2014) perform a similar analysis for a dichotomy in which voters have independent (but not identically distributed) random utilities. They show that the rule which maximizes *ex ante* utilitarian social welfare over the class of all incentive compatible rules is a weighted majoritarian rule (where the weight of each voter is determined by the expected value of her utility function). They also obtain a similar characterization of the *ex ante* and *ex interim* Pareto-optimal rules in the class of incentive-compatible rules.

The majority of the aforementioned papers deal only with dichotomous decisions, whereas we allow an arbitrary number of alternatives.¹⁰ Also, except for Fleurbaey (2009), all of the aforementioned papers assumed that cardinal interpersonal utility comparisons are unproblematic. In contrast, we suppose that these interpersonal comparisons themselves are ambiguous in practice (but still meaningful in principle). Finally, except for Weber (1978) and Caragiannis and Procaccia (2011), all of the aforementioned papers deal with “small” populations of voters, whereas we are interested in asymptotic probabilistic

⁹For i.u.d. utilities, this result had been anticipated by Weber (1978a, p.10).

¹⁰However, approval voting and all rank scoring rules reduce to simple majority vote when there are only two alternatives. Thus, in this setting, our results in Sections 3 and 4 imply the utilitarian optimality of simple majority voting, and are complementary to the Rae-Taylor theorem and its extensions.

results for very large populations.¹¹ It is not possible here to adequately summarise the vast and growing literature on the large-population asymptotic probabilistic analysis of voting rules. Instead, we will only briefly touch on two strands of this literature. The first strand is the Condorcet Jury Theorem (CJT) and its many generalizations.¹² Like the CJT literature, the results of the present paper say that, under certain probabilistic assumptions, a large population using a certain voting rule is likely to arrive at the “correct” decision. But the goal for the CJT literature is to find the correct answer to some objective factual question, whereas the goal in the present paper is to maximize social welfare.

The second strand is the literature on strategic voting and/or strategic candidacy in large populations with some kind of randomness or uncertainty in voters’ preferences. This literature is mainly concerned with characterizing the Nash equilibria of certain large election games. These equilibria occasionally have surprising social welfare properties. For example, Ledyard (1984), Lindbeck and Weibull (1987, 1993), Coughlin (1992; Theorem 3.7 and Corollary 4.4), Banks and Duggan (2004; §4) and McKelvey and Patty (2006) have all shown that, in certain election games, there is a unique Nash equilibrium (sometimes called a “political equilibrium”) where all the candidates select the policy which maximizes a utilitarian SWF. But these utilitarian SWFs are based on somewhat peculiar systems of interpersonal utility comparisons. In these models, voter behaviour is described by a stochastic device: the probability that voter i votes for candidate C (or in some cases, the probability that i votes at all) is a function of the difference between the cardinal utility which i assigns to C and the cardinal utility she assigns to other candidates. Although the different models use different stochastic devices and seek to capture different phenomena (e.g. random private costs for voting, or random private shocks to the utility functions, or random individual errors due to bounded rationality, or other exogenous perturbations), each model assumes that utility functions are translated into voting probabilities *in the same way* for every voter. In this way, each model smuggles in a system of “implicit” interpersonal utility comparisons via the stochastic device. As observed by Banks and Duggan (2004, p.29), this means that the normative significance of the “utilitarianism” emerging from these political equilibria is somewhat unclear.

In contrast, this paper assumes that there is a pre-existing, normatively meaningful system of cardinal interpersonal utility comparisons, explicitly described by a set of “calibration constants” which exist independently of the voting rule and any other random factors in the model. The social planner does not know the exact values of these calibration constants, so she regards them as random variables. Our results suggest that it is still possible to closely approximate the utilitarian social choice, even with this kind of uncertainty. On the other hand, unlike the political equilibrium literature described in the previous paragraph, this paper does not grapple with strategic issues, except in the conclusion. Also, unlike the political equilibrium literature, this paper treats the social al-

¹¹Since they consider a democratic federation of states, the papers by Barberà and Jackson (2006), Beisbart, Bovens and Hartmann, etc. presumably posit large populations. However, most of these papers represent each population in reduced form as an averaged utility function, not as a set of individuals, and none of them engage in any sort of asymptotic analysis.

¹²See Nitzan (2009, Ch.11-12) or Pivato (2013) for surveys of this literature.

ternatives as exogenous, rather than endogenizing them as the result of political candidates competing for popularity.

2 Evaluative voting

The most natural “utilitarian” voting rule simply asks the each voter to assign a numerical score to each alternative, presumably reflecting the utility she would obtain from that alternative. The obvious problem with this approach is that voters could strategically exaggerate their scores. One partial solution is to rescale every voter’s utility function to range over the interval $[0, 1]$.¹³ The resulting social choice rule has been called *evaluative voting* (Núñez and Laslier, 2014), *utilitarian voting* (Hillinger, 2005), or *range voting* (Smith, 2000; Macé, 2013). It is closely related to the *relative utilitarian* SWF axiomatized by Dhillon (1998) and Dhillon and Mertens (1999).

Before continuing, let us fix some notation which will be maintained throughout the paper. Let \mathbb{R} denote the set of real numbers. For any set \mathcal{A} and function $f : \mathcal{A} \rightarrow \mathbb{R}$, let $\operatorname{argmax}_{\mathcal{A}}(f)$ denote the set of elements in \mathcal{A} which maximize f . For any random variable X , let $\mathbb{E}[X]$ denote its expected value, and let $\operatorname{var}[X]$ denote its variance.

Let \mathcal{A} be a finite set of social alternatives, let \mathcal{I} be a set of voters, and let $I := |\mathcal{I}|$. (We will typically suppose that I is very large.) In *evaluative voting* (EV), vote of each voter i in \mathcal{I} takes the form of a function $v_i : \mathcal{A} \rightarrow [0, 1]$. The EV rule then chooses the alternative(s) in \mathcal{A} that maximize the function $V_{\mathcal{I}} : \mathcal{A} \rightarrow \mathbb{R}$ defined by

$$V_{\mathcal{I}}(a) := \sum_{i \in \mathcal{I}} v_i(a), \quad \text{for every alternative } a \text{ in } \mathcal{A}. \quad (1)$$

Does EV also maximize the utilitarian social welfare function? If not, how close does it come to this goal? For every voter i in \mathcal{I} let $w_i : \mathcal{A} \rightarrow \mathbb{R}$ be her “true” utility function. We suppose these utility functions admit one-for-one cardinal interpersonal comparisons. In other words, for any alternatives a, b, c , and d in \mathcal{A} , if $w_i(b) - w_i(a) = w_j(d) - w_j(c)$, then the welfare that voter i gains in moving from a to b is exactly the same as the welfare that voter j gains in moving from c to d . We therefore want to maximize the utilitarian SWF $U_{\mathcal{I}}$ defined by

$$U_{\mathcal{I}}(a) := \frac{1}{I} \sum_{i \in \mathcal{I}} w_i(a), \quad \text{for every alternative } a \text{ in } \mathcal{A}. \quad (2)$$

Let $\underline{w}_i := \min\{w_i(a); a \in \mathcal{A}\}$. By replacing w_i with the function $\tilde{w}_i := w_i - \underline{w}_i$ if necessary, we can suppose that $\min\{w_i(a); a \in \mathcal{A}\} = 0$, for every voter i in \mathcal{I} . Clearly this does not affect the maximizer of (2).

Next, let $c_i := \max\{w_i(a); a \in \mathcal{A}\}$, and then define $u_i(a) := w_i(a)/c_i$, for every voter i in \mathcal{I} and every alternative a in \mathcal{A} . Note that each u_i ranges over the interval $[0, 1]$. The *relative utilitarian* social welfare function $RU : \mathcal{A} \rightarrow \mathbb{R}$ is defined:

$$RU_{\mathcal{I}}(a) := \sum_{i \in \mathcal{I}} u_i(a), \quad \text{for every alternative } a \text{ in } \mathcal{A}. \quad (3)$$

¹³Obviously, this is not a complete solution to strategic voting.

The scaling constants $\{c_i\}_{i \in \mathcal{I}}$ represent the “preference intensities” of the voters, which we assume are unknown to the social planner. From the planner’s point of view, $\{c_i\}_{i \in \mathcal{I}}$ are random variables. We formalize this with the following assumption:

- (E1)** $\{c_i\}_{i \in \mathcal{I}}$ are real-valued random variables, which are independent, but not necessarily identically distributed. There is some constant $\sigma_c^2 \geq 0$ such that $\text{var}[c_i] \leq \sigma_c^2$ and $\mathbb{E}[c_i] = 1$ for all $i \in \mathcal{I}$.

Hypotheses (E1) does not allow the preference intensities of different voters to be correlated; it cannot accommodate a scenario where voter i is likely to have strong preferences if voter j does. However, (E1) *does* allow that certain types of voters tend to have more intense preferences than other voters, for certain types of policy problems; for example, parents of young children might have especially intense preferences about education policies, while industrialists might have strong preferences about regulations affecting their industry.

Assuming voter i uses the full range $[0, 1]$ to express her utilities, but is otherwise accurate, she will set $v_i = w_i$. If all voters behave in this way, then $V_{\mathcal{I}} = RU_{\mathcal{I}}$. But voter i may misperceive her own utility function. Thus, in general, $v_i = u_i + \epsilon_i$, where $\epsilon_i : \mathcal{A} \rightarrow \mathbb{R}$ is a random “error” function. Suppose \underline{a}_i and \bar{a}_i are the minimizer and maximizer of u_i (thus, $u_i(\underline{a}_i) = 0$ and $u_i(\bar{a}_i) = 1$). It is reasonable to suppose that $v_i(\underline{a}_i) = 0$ and $v_i(\bar{a}_i) = 1$ —that is, voter i reliably assigns a score of 0 to her worst alternative and a score of 1 to her best alternative. Thus, $\epsilon_i(\underline{a}_i) = \epsilon_i(\bar{a}_i) = 0$. However, for the other alternatives in \mathcal{A} , the errors may be nonzero. We assume they satisfy the following condition:

- (E2)** For each alternative a in \mathcal{A} , the random errors $\{\epsilon_i(a)\}_{i \in \mathcal{I}}$ are independent,¹⁴ but not necessarily identically distributed. There is some constant $\sigma_\epsilon^2 > 0$ such that $\text{var}[\epsilon_i(a)] \leq \sigma_\epsilon^2$ and $\mathbb{E}[\epsilon_i(a)] = 0$ for all $a \in \mathcal{A}$ and $i \in \mathcal{I}$.¹⁵ The random variables $\{c_i\}_{i \in \mathcal{I}}$ are independent of the random functions $\{\epsilon_i\}_{i \in \mathcal{I}}$.

Finally, we assume the utility profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies the following technical property.

- (E3)** There is a constant $\Delta > 0$ such that $\max_{\mathcal{A}}(U_{\mathcal{I}}) - U_{\mathcal{I}}(a) > \Delta$ for every $a \notin \text{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$.¹⁶

Here, Δ is the minimum social welfare gap between the optimal policy and the next-best policy. If a voting rule acts as an “estimator” of the utilitarian SWF, then we need the error of this estimate to be smaller than Δ , in order for the rule to select the optimal policy, and not the next-best policy.¹⁷ Our first result says that, despite the uncertainties surrounding $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i\}_{i \in \mathcal{I}}$, evaluative voting has a very good chance of maximizing the utilitarian social welfare function $U_{\mathcal{I}}$ when the population is large.

¹⁴Note that we do *not* assume that, for a fixed voter i in \mathcal{I} , the random errors $\epsilon_i(a)$ and $\epsilon_i(b)$ are independent for different alternatives a and b in \mathcal{A} .

¹⁵This is consistent with the fact that $\epsilon_i(\underline{a}_i) = \epsilon_i(\bar{a}_i) = 0$ with certainty —this just means that the distributions of $\epsilon_i(\underline{a}_i)$ and $\epsilon_i(\bar{a}_i)$ are point masses.

¹⁶Most of our results involve taking a limit as $I \rightarrow \infty$. In these results, Δ is to be held constant as $I \rightarrow \infty$.

¹⁷Of course, if Δ was very small, then selecting the next-best policy would not be a catastrophe; thus, we will relax condition (E3) in Proposition 2 below.

Theorem 1 *For every voter i in \mathcal{I} , let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be a utility function. Suppose that the profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies (E3), and suppose $\{c_i\}_{i \in \mathcal{I}}$, $\{\epsilon_i\}_{i \in \mathcal{I}}$ and $\{v_i\}_{i \in \mathcal{I}}$ are randomly generated according to rules (E1) and (E2). Define $V_{\mathcal{I}}$ and $U_{\mathcal{I}}$ as in equations (1) and (2).*

Then $\lim_{I \rightarrow \infty} \text{Prob} \left[\underset{\mathcal{A}}{\text{argmax}}(V_{\mathcal{I}}) \subseteq \underset{\mathcal{A}}{\text{argmax}}(U_{\mathcal{I}}) \right] = 1$.

We can refine this result in three ways. First, we can drop condition (E3). Second, and relatedly, instead of demanding that the outcome of evaluative voting *exactly* maximizes $U_{\mathcal{I}}$, we can allow the possibility that it only *almost* maximizes $U_{\mathcal{I}}$ —something which would be almost as good, for practical purposes. Third, we can estimate how *large* the population needs to be in order to achieve such “almost-maximization” with a certain probability. To achieve these refinements, we need some more notation. For any utility profile $\{u_i\}_{i \in \mathcal{I}}$, if $U_{\mathcal{I}}$ is as in equation (2), then let

$$U_{\mathcal{I}}^* := \max \{U_{\mathcal{I}}(a) ; a \in \mathcal{A}\} \quad (4)$$

This is the theoretical maximum social welfare, which would be obtained from the optimal social alternative. Let $\delta > 0$ represent a “social suboptimality tolerance”, and let $p > 0$ represent the probability that this tolerance will be exceeded (the social planner wants both of these to be small). For any values of δ and p , we define

$$\bar{I}(\delta, p) := 4A \frac{\sigma_c^2 + \sigma_\epsilon^2}{p \delta^2}. \quad (5)$$

Our next result says that, for any population larger than $\bar{I}(\delta, p)$, any $V_{\mathcal{I}}$ -maximizing social alternative will produce a utilitarian social welfare within δ of the theoretical optimum, with probability at least $1 - p$.

Proposition 2 *For every voter i in \mathcal{I} , let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be a utility function. Suppose $\{c_i\}_{i \in \mathcal{I}}$, $\{\epsilon_i\}_{i \in \mathcal{I}}$ and $\{v_i\}_{i \in \mathcal{I}}$ satisfy (E1) and (E2). Define $V_{\mathcal{I}}$, $U_{\mathcal{I}}$, and $U_{\mathcal{I}}^*$ as in equations (1), (2), and (4). For any $\delta > 0$ and $p \in (0, 1)$, if $I \geq \bar{I}(\delta, p)$, then $\text{Prob} [U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta] < p$, for every a in $\underset{\mathcal{A}}{\text{argmax}}(V_{\mathcal{I}})$.*

The evaluative voting rule prevents voters from strategically exaggerating their utility functions, but strategic voting is still possible. In general, each voter’s best response is to assign a score of either 0 or 1 to each alternative in \mathcal{A} (Núñez and Laslier, 2014). In this case, evaluative voting reduces to approval voting.

3 Approval voting

*Approval voting*¹⁸ works as follows:

¹⁸Approval voting was originally proposed by Ottowell (1977), Kellett and Mott (1977), and Weber (1978), but the first sustained formal analysis was in Brams and Fishburn (1983), which has become the standard reference on the subject.

1. Each voter i identifies a subset of alternatives in \mathcal{A} which she “approves”.
2. For each social alternative a in \mathcal{A} , count how many voters approve a .
3. Choose the alternative which is approved by the most voters.

Formally, for every voter i in \mathcal{I} , let $\mathcal{G}_i \subseteq \mathcal{A}$ be the set of alternatives which i approves; we refer to \mathcal{G}_i as her *approval set*. Let $\mathcal{G} := \{\mathcal{G}_i\}_{i \in \mathcal{I}}$ be the profile of the voters’ approval sets. For any social alternative a in \mathcal{A} , we define its *approval score* by: $V_{\mathcal{G}}(a) := \#\{i \in \mathcal{I}; a \in \mathcal{G}_i\}$. We then define $\text{Appr}(\mathcal{G}) := \text{argmax}_{\mathcal{A}}(V_{\mathcal{G}})$.

For every i in \mathcal{I} , let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be i ’s cardinal utility function, and let $c_i > 0$ be a “calibration constant”, which we will use to make cardinal interpersonal utility comparisons. We suppose that the functions $c_i u_i$ and $c_j u_j$ are interpersonally comparable for all voters i and j in \mathcal{I} . In other words, for any alternatives a, b, c , and d in \mathcal{A} , if $c_i u_i(b) - c_i u_i(a) = c_j u_j(d) - c_j u_j(c)$, then the welfare that voter i gains in moving from alternative a to alternative b is exactly the same as the welfare that voter j gains in moving from c to d . We would therefore like to maximize the utilitarian social welfare function $U_{\mathcal{I}} : \mathcal{A} \rightarrow \mathbb{R}$ defined by

$$U_{\mathcal{I}}(a) := \frac{1}{I} \sum_{i \in \mathcal{I}} c_i u_i(a), \quad \text{for every alternative } a \text{ in } \mathcal{A}. \quad (6)$$

If the approval set \mathcal{G}_i is a “noisy signal” of u_i (for every voter i in \mathcal{I}), then the aggregate approval score $V_{\mathcal{G}}$ could be seen as a “noisy signal” of the social welfare function $U_{\mathcal{I}}$. Thus, under the right conditions, approval voting should maximize utilitarian social welfare. Weber (1978) showed this was true when voters have i.u.d. utility functions over a set of three alternatives. The goal of this section is to make this intuition precise in a much more general setting.

Each voter’s true utility function is unknown to the social planner. So is the process by which each voter converts her utility function into an approval set. The planner deals with this uncertainty by treating these as random variables, described by some probabilistic model. We will consider two different models: the Threshold Model and the Selection Model. The Threshold Model first assigns each voter a random utility for each social alternative, and then selects her approval set from these alternatives by means of a randomly determined threshold. The Selection Model first selects an approval set for each voter (this process may be random or deterministic), and then randomly assigns utilities to each social alternative, according to a probability distribution which depends on whether or not it is in the approval set. Both models yield the same conclusion: in a large population, approval voting maximizes the utilitarian social welfare function, with high probability.

3.1 The Threshold Model

For each voter i in \mathcal{I} , and every alternative a in \mathcal{A} , let u_a^i be the utility which i assigns to a . We suppose that i also identifies an *approval threshold* θ_i . She then defines her approval

set \mathcal{G}_i to be all social alternatives whose utility exceeds θ_i .¹⁹ That is:

$$\mathcal{G}_i := \{a \in \mathcal{A}; u_a^i \geq \theta_i\}. \quad (7)$$

We make the following assumptions:

- (Θ1) Same as (E1).
- (Θ2) The utilities $\{u_a^i; a \in \mathcal{A} \text{ and } i \in \mathcal{I}\}$ are i.i.d. random variables with finite variance. The variables $\{c_i\}_{i \in \mathcal{I}}$ and $\{u_i\}_{i \in \mathcal{I}}$ are all independent.
- (Θ3) Same as (E3).

Finally, we assume:

- (Θ0) The thresholds $\{\theta_i\}_{i \in \mathcal{I}}$ are independent random variables (not necessarily identically distributed). For any $i \in \mathcal{I}$, the random threshold θ_i may depend on $\{u_a^i\}_{a \in \mathcal{A}}$. But $\{\theta_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ are independent. Finally, $0 < \text{Prob}[u_a^i \geq \theta_i] < 1$ for all $a \in \mathcal{A}$.

Assumptions (Θ1) and (Θ3) have the same interpretations as (E1) and (E3). Assumption (Θ2) describes the planner’s ignorance about people’s true utility functions, while (Θ0) describes her ignorance about how they convert these utility functions into approval sets, except for the fact that they follow rule (7). The condition “ $0 < \text{Prob}[u_a^i \geq \theta_i] < 1$ ” simply guarantees that the output of rule (7) is almost-surely nondegenerate. The next result says that, if a large population of voters satisfies hypotheses (Θ0)-(Θ3), then with very high probability, approval voting will maximize the utilitarian social welfare function $U_{\mathcal{I}}$ in equation (6).

Theorem 3 *Suppose $\{u_i\}_{i \in \mathcal{I}}$, $\{\theta_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (Θ0)-(Θ3), and the approval profile $\mathcal{G} = \{\mathcal{G}_i\}_{i \in \mathcal{I}}$ is defined by rule (7). Then $\lim_{I \rightarrow \infty} \text{Prob} \left[\text{Appr}(\mathcal{G}) \subseteq \underset{\mathcal{A}}{\text{argmax}}(U_{\mathcal{I}}) \right] = 1$.*

One notable difference between Theorem 1 and Theorem 3 is that the former places essentially no conditions on the utility functions $\{u_i\}_{i \in \mathcal{I}}$, whereas the latter requires these utility functions to be i.i.d. random variables. The reason is that approval voting provides us with less information about individual utility functions than evaluative voting. A voter’s response in evaluative voting tells us her entire utility function, up to a scalar multiple; the only uncertainty is the magnitude of this scalar. But her approval set only tells us whether each she assigns a ‘high’ or ‘low’ utility to each alternative. Thus, without further information about the conditional probability distributions of these ‘high’ and ‘low’ utilities, it is not possible for us to estimate the utilitarian social welfare function.

Note that the total ignorance described by (Θ2) implies both a sort of *a priori anonymity* (i.e. all voters are indistinguishable, *a priori*) and *a priori neutrality* (i.e. all social alternatives are indistinguishable, *a priori*). Thus, it does not allow us to incorporate the

¹⁹In the theory of approval voting, this is known as a *sincere* voting strategy. It is easy to see that any insincere voting strategy is weakly dominated by some sincere voting strategy.

knowledge that certain alternatives (e.g. low taxes) tend to be favoured by certain classes of voters (e.g. business owners). Nor does it allow us to incorporate the knowledge that people’s preferences may be correlated (e.g. voters who favour low taxes tend to also favour less regulation, because they are often business owners). Likewise, $(\Theta 0)$ does not allow us to incorporate knowledge that certain types of voters tend to set higher thresholds than others for certain types of policy problems. (However, $(\Theta 1)$, like $(E1)$, does allow certain types of voters to tend towards more intense preferences than others.) Thus, when applied to a specific policy problem, these assumptions may be less than optimal; a voting rule more closely optimized to the specific probability distribution of utility functions in a society might yield a higher expected social welfare.²⁰

However, in many cases, we may not have actionable information about the distribution of utility functions. Furthermore, at a political level, it could be problematic to deploy a “customized” voting rule for each policy issue; it would be too easy for such “customization” to smuggle in the biases of the social planner, either deliberately or inadvertently. Even if these customized voting rules were developed through some impeccably impartial process, based on objective and publicly known facts about the statistical distribution of voter preferences, there might remain the suspicion that the outcome was due to some manipulation of this process, and this could undermine its legitimacy. If this is a real concern, then it is important to fix a single voting rule once and for all when designing the constitution, and then apply this one voting rule to all policy questions. In the design of such a constitution, the *a priori* anonymity and neutrality of $(\Theta 0)$ and $(\Theta 2)$ is entirely appropriate.

3.2 The Selection Model

In our second model, the approval set of each voter is exogenously and arbitrary. Her approval set might be fixed in advance, or it might be generated by some other random process. (If the voters’ approval sets *are* random variables, then we do not need to assume that they are either independent, or identically distributed.) Each voter assigns random utilities to each alternative, conditional on whether or not it is in her approval set.

Formally, for each i in \mathcal{I} , let \mathcal{G}_i be the (exogenous) approval set of voter i , and let $\mathcal{B}_i := \mathcal{A} \setminus \mathcal{G}_i$. Let γ and β be two finite-variance probability measures on \mathbb{R} ,²¹ such that the mean value of γ is strictly larger than that of β . We make the following assumptions:

(S1) Same as $(\Theta 1)$ and $(E1)$.

(S2) For all g in \mathcal{G}_i , $u_i(g)$ is a γ -random variable. For all b in \mathcal{B}_i , $u_i(b)$ is a β -random variable.²² The random variables $\{c_i; i \in \mathcal{I}\}$ and $\{u_i(a); i \in \mathcal{I} \text{ and } a \in \mathcal{A}\}$ are all

²⁰For example, the scoring rules of Kim (2014;§5) and the weighted majority rules of Bordley (1985b, 1986), Fleurbaey (2009) and Azrieli and Kim (2014;§4) have this feature.

²¹These are not necessarily Gamma or Beta distributions.

²²If the supports of γ and β overlap, then there is a chance that the voter will assign a *lower* utility to some element of \mathcal{G}_i than she assigns to some element of \mathcal{B}_i . This would appear somewhat irrational. To exclude this, one could assume that γ is a finite-variance probability measure on $(0, \infty)$, while β is a finite-variance probability measure on $(-\infty, 0]$. But our results do not require this assumption.

independent.

(S3) Same as (Θ3) and (E3).

The interpretation of (S1) is the same as (E1) or (Θ1). The interpretation of (S2) is the same as (Θ2); it implies both a sort of *a priori anonymity* and *a priori neutrality*. The difference is that now we can distinguish, *a priori* between those alternatives which *are* in a voter’s approval set and those which aren’t, and give them different statistical treatments. Since the approval sets themselves are exogenous, this model is perfectly able to cope with situations where a voter’s approval choices are highly correlated, both with other voters and with other known facts about that voter (e.g. the fact that business owners tend to approve of tax reduction and *also* tend to approve of deregulation). The model in effect makes *no* assumptions about the statistical distribution of approval sets (it doesn’t even treat them as random). However, the model still does not allow correlations of utility *within* each voter’s approval set (e.g. a correlation between the utility that a voter assigns to a tax reduction and the utility she assigns to a deregulation policy, given that she has approved of both). The earlier remarks about constitutional design still apply. Once again, if a large population of voters satisfies these hypotheses, then with very high probability, approval voting will maximize the utilitarian social welfare function $U_{\mathcal{I}}$ in equation (6).

Theorem 4 *Let $\mathcal{G} = \{\mathcal{G}_i\}_{i \in \mathcal{I}}$ be an arbitrary approval profile. If $\{u_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ satisfy hypotheses (S1)-(S3), then $\lim_{I \rightarrow \infty} \text{Prob} \left[\text{Appr}(\mathcal{G}) \subseteq \underset{\mathcal{A}}{\text{argmax}}(U_{\mathcal{I}}) \right] = 1$.*

As in Section 2, we would like to refine Theorems 3 and 4 by dropping conditions (Θ3) and (S3). We would also like to estimate how large the population must be in order for approval voting to “almost-maximize” $U_{\mathcal{I}}$ with a certain probability, by analogy with Proposition 2. For brevity, we will present such a result only for the Selection Model, but a similar result can be proved for the Threshold Model. Let $U_{\mathcal{I}}^* := \max\{U_{\mathcal{I}}(a); a \in \mathcal{A}\}$.

Proposition 5 *Let $\mathcal{G} = \{\mathcal{G}_i\}_{i \in \mathcal{I}}$ be an arbitrary approval profile. If $\{u_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (S1) and (S2), then for any $\delta > 0$, we have*

$$\lim_{I \rightarrow \infty} \text{Prob} \left(U_{\mathcal{I}}(a) \geq U_{\mathcal{I}}^* - \delta \text{ for all } a \in \text{Appr}(\mathcal{G}) \right) = 1. \quad (8)$$

Furthermore, if the fourth moments of γ and β are finite, then there are constants $C_1, C_2 > 0$ (determined by γ, β , and σ_c^2) such that, for any $p > 0$, if $I \geq C_1/p$ and $I \geq C_2/p\delta^2$, then $\text{Prob}[U_{\mathcal{I}}(a) < U_{\mathcal{I}}^ - \delta] < p$ for all $a \in \text{Appr}(\mathcal{G})$.*

4 Rank scoring rules

One concern with approval voting is that it gives each voter very little ability to express the intensity of her preference for or against each alternative. For example, if a voter does not include a certain alternative in her approval set, this may be because she actively

dislikes this alternative, or it may simply be because she has no strong feelings either way, or perhaps inadequate knowledge of this alternative, and for this reason she “abstains” from endorsing it. Approval voting is unable to distinguish between voting against and abstention. For this reason, Alcantud and Laruelle (2014) propose “dis&approval” voting, which allows a voter three choices for each alternative: approve, disapprove, or abstain. However, “dis&approval” voting still does not distinguish between “strong” (dis)approval and “weak” (dis)approval. For this purpose, we turn to rank scoring rules.

Let $N := \mathcal{A}$. Let $s_1 \leq s_2 \leq \dots \leq s_N$ be real numbers, and define $\mathbf{s} := (s_1, s_2, \dots, s_N)$. The \mathbf{s} -rank scoring rule on \mathcal{A} is defined as follows:

1. For every voter i in \mathcal{I} , let \succ_i denote her (strict) ordinal preferences on \mathcal{A} .
2. For every alternative a in \mathcal{A} , if a is ranked k th place from the bottom with respect to \succ_i , then voter i gives a the score s_k . (In particular, i gives the score s_1 to her least-preferred alternative, and the score s_N to her most preferred alternative.)
3. For each alternative in \mathcal{A} , add up the scores it gets from all voters.
4. The \mathbf{s} -rank scoring rule chooses the alternative(s) with the highest total score.

For example, the *Borda rule* is the rank scoring rule with $\mathbf{s} = (1, 2, 3, \dots, N)$. The standard *plurality rule* is the rank scoring rule with $\mathbf{s} = (0, 0, \dots, 0, 1)$.

Formally, for every voter i in \mathcal{I} , if $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$ and $a_1 \prec_i a_2 \prec_i \dots \prec_i a_N$, then define $v_i : \mathcal{A} \rightarrow \mathbb{R}$ by setting $v_i(a_k) := s_k$ for all $k \in [1 \dots N]$. Let $\mathcal{P} := \{\succ_i\}_{i \in \mathcal{I}}$ be the profile of ordinal preferences of the voters. For every social alternative a in \mathcal{A} , define $V_{\mathcal{P}}^{\mathbf{s}}(a) := \sum_{i \in \mathcal{I}} v_i(a)$. Then define $\text{Score}_{\mathbf{s}}(\mathcal{P}) := \text{argmax}_{\mathcal{A}}(V_{\mathcal{P}}^{\mathbf{s}})$.

Recently, Apesteguia et al. (2011; Theorem 3.1) have shown that, amongst all voting rules, rank scoring rules are the ones which maximize the *expected value* of the utilitarian social welfare function (under certain conditions). Furthermore, they characterized the optimal rank scoring rule in terms of the probability distribution of the voters’ utility functions—to be precise, in terms of the *expected order statistics* of this distribution. Our results in this section are complementary. We work with a much broader class of probability distributions than Apesteguia et al. (2011). We will show that, if the profile $\{u_i\}_{i \in \mathcal{I}}$ arises from this class, then there exists a rank scoring rule which will come arbitrarily close to selecting a utilitarian optimum, with very high probability as $I \rightarrow \infty$. Thus, while Apesteguia et al. (2011) show that the optimal rank scoring rule is “better on average” than any other voting rule, we show that it is, in fact “almost perfect”, in the limit of a large population.²³

As in Section 3, we will present two stochastic models of voter preference formation. In the *Endogenous Preference* model, the voters’ utility functions are i.i.d. random variables, and their ordinal preferences are determined by these utility functions. In the *Exogenous preference* model, the voter’s ordinal preferences are exogenous and arbitrary, and their utility functions are random variables conditional on these preferences.

²³Weber (1978) showed this was true for the Borda rule, when voters have i.u.d. utility functions.

4.1 Endogenous Preferences

For each voter i in \mathcal{I} , we will represent her utility function over \mathcal{A} as an N -dimensional vector $\mathbf{u}^i = (u_a^i)_{a \in \mathcal{A}} \in \mathbb{R}^{\mathcal{A}}$. Let μ be a finite-variance²⁴ probability measure on $\mathbb{R}^{\mathcal{A}}$; we will use this probability measure to randomly generate the utility functions of the voters. For any $a, b \in \mathcal{A}$, we assume that $\mu\{\mathbf{u} \in \mathbb{R}^{\mathcal{A}}; u_a = u_b\} = 0$ (i.e. almost surely, no two alternatives yield the same utility). We also make the following assumptions.

(N0) μ is symmetric under all coordinate permutations.

(N1) Same as (S1), (Θ 1), and (E1).

(N2) The utility functions $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$ are independent, μ -random vectors. Also, $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$ are independent of $\{c_i\}_{i \in \mathcal{I}}$.

(N3) Same as (S3), (Θ 3), and (E3).

Note that we do *not* assume, for any particular voter i , that the utilities she assigns to different alternatives a and b are independent random variables: the probability measure μ may allow for arbitrary correlations between u_a^i and u_b^i . However, assumption (N0) acts as a form of *a priori neutrality*. It implies that the expected value of u_a^i and u_b^i are the same. Furthermore, the expected value of u_a^i , conditional on some information about the rank of b (say, that b is i 's favourite alternative) is the same as the expected value of u_b^i conditional on the same information about the rank of a . Hypothesis (N1) is comparable to (S1) or (Θ 1). Hypothesis (N2) is comparable to (S2) or (Θ 2): it is a sort of *a priori anonymity*, saying that all voters are indistinguishable, *a priori*. However, (N2) supposes a richer level of knowledge than (S2) or (Θ 2); for example, we might know that, on average, each voter's second-best and third-best alternatives obtain, respectively 90% and 75% of the utility of her favourite alternative (assuming her worst alternative has utility 0). We will now see how to incorporate such knowledge into the optimal rank scoring rule.

Let $\mathbf{u} \in \mathbb{R}^{\mathcal{A}}$ be a μ -random variable. Rearrange the coordinates of \mathbf{u} in increasing order, to get a vector $\hat{\mathbf{u}} := (\hat{u}_1, \dots, \hat{u}_N)$, where $\hat{u}_1 \leq \hat{u}_2 \leq \dots \leq \hat{u}_N$. For all $n \in [1 \dots N]$, let s_n be the expected value of \hat{u}_n ; this yields a vector $\mathbf{s} = (s_1, \dots, s_N)$. The next result says that, if a large population of voters satisfies hypotheses (N0)-(N3), then with very high probability, the \mathbf{s} -rank scoring rule will maximize the utilitarian social welfare function $U_{\mathcal{I}}$ in equation (6).

Theorem 6 *Suppose $\{u_i\}_{i \in \mathcal{I}}$, and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (N0)-(N3), and let $\mathcal{P} = \{\succsim_i\}_{i \in \mathcal{I}}$ be the ordinal preference profile defined by $\{u_i\}_{i \in \mathcal{I}}$.²⁵ Then*

$$\lim_{I \rightarrow \infty} \text{Prob} \left[\text{Score}_{\mathbf{s}}(\mathcal{P}) \subseteq \underset{\mathcal{A}}{\text{argmax}}(U_{\mathcal{I}}) \right] = 1. \quad (9)$$

²⁴By this, we mean that the variance of each coordinate of a μ -random variable is finite.

²⁵ \mathcal{P} is almost surely a profile of *strict* preferences, because by hypothesis on μ , no two alternatives yield the same utility for any voter, almost surely.

Impartial Culture. As we have emphasized, hypothesis (N2) does *not* require the utilities $\{u_a^i\}_{a \in \mathcal{A}}$ to be i.i.d. random variables, for any particular voter i . However, it is certainly compatible with this additional assumption. This corresponds to the special case of the model where μ is a Cartesian product of N copies of some underlying finite-variance, nonatomic probability measure ρ on \mathbb{R} . In this case, the utilities $\{u_a^i; i \in \mathcal{I} \text{ and } a \in \mathcal{A}\}$ are all independent, ρ -random variables (this is a version of the so-called “Impartial Culture” model). If we make the further assumption that ρ is absolutely continuous and has compact support, and assume that $c_i := 1$ for every voter i in \mathcal{I} , then we obtain the models considered by Weber (1978) and Apesteguia et al. (2011). But the Endogenous Preference model is much more general than this Impartial Culture model.²⁶

4.2 Exogenous Preferences

In our second model, the preference orders of the voters are exogenous and arbitrary. These preference orders may themselves be random variables, or they may be determined in some other way. (If they *are* random variables, then we do not assume that they are independent or identically distributed. In particular, unlike the Impartial Culture model, we do *not* suppose that all $N!$ possible ordinal preferences on \mathcal{A} are equally likely to occur in \mathcal{P} .) The model then assigns each voter a random utility for each alternative, conditional on her preference order.

Formally, let $\mathbb{U} := \{\mathbf{u} \in \mathbb{R}^N; u_1 < u_2 < \dots < u_N\}$, and let λ be a finite-variance probability measure on \mathbb{U} . For every voter i in \mathcal{I} , let \succ_i denote i 's (exogenous) preference order over \mathcal{A} . We make the following assumptions:

(X1) Same as (N1), (S1), (Θ 1) and (E1).

(X2) The utility vectors $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$ are independent random vectors, generated as follows. For each $i \in \mathcal{I}$, we obtain \mathbf{u}^i by taking a λ -random variable $\widehat{\mathbf{u}}^i$, and rearranging the coordinates to agree with the preference order \succ_i . The random variables $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$ and $\{c^i\}_{i \in \mathcal{I}}$ are independent.

(X3) Same as (N3), (S3), (Θ 3), and (E3).

Once again, we do *not* assume, for any particular voter i , that the utilities she assigns to different alternatives a and b are independent random variables, even after we condition on \succ_i ; the probability measure λ may allow for arbitrary correlations between u_a^i and u_b^i . However, (X2) has a consequence similar to (N0): given two voters i and j , and two alternatives a and b , if a has the same rank with respect to \succ_i as b does with respect to \succ_j , then the expected value of u_a^i is the same as that of u_b^j . The interpretation of hypothesis (X1) is the same as to (N1). Hypothesis (X2) is comparable to (N2): it is a sort of *a priori*

²⁶One problem with Impartial Culture models is that, in a large population, all elements of \mathcal{A} end up with roughly the *same* average utility (due to the Law of Large Numbers), so that utilitarianism is effectively indifferent between them, and the use of *any* voting rule is somewhat superfluous. The Endogenous Preference model avoids this unrealistic outcome.

anonymity, saying that all voters are indistinguishable, *a priori*, except for their exogenous ordinal preferences.

For all $k \in [1 \dots N]$, let s_k be the expected value of u_k , where $(u_1, \dots, u_N) \in \mathbb{U}$ is a λ -random variable. Let $\mathbf{s} := (s_1, s_2, \dots, s_N)$. The next result says that, if a large population of voters satisfies hypotheses (X1)-(X3), then with very high probability, the \mathbf{s} -rank scoring rule will maximize the utilitarian social welfare function $U_{\mathcal{I}}$ in equation (6).

Theorem 7 *Let $\mathcal{P} = \{\succ_i\}_{i \in \mathcal{I}}$ be an arbitrary ordinal preference profile. Suppose $\{u_i\}_{i \in \mathcal{I}}$, and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (X1)-(X3). Then the limit (9) holds.*

As in Sections 2 and 3, we would like to refine Theorems 6 and 7 by dropping conditions (N3) and (X3). We would also like to estimate how large the population must be in order for the rank scoring rule to “almost-maximize” $U_{\mathcal{I}}$ with a certain probability, by analogy with Propositions 2 and 5. For brevity, we will present such a result only for the Exogenous Preference model, but a similar result can be proved for the Endogenous Preference model. As usual, let $U_{\mathcal{I}}^* := \max\{U_{\mathcal{I}}(a); a \in \mathcal{A}\}$.

Proposition 8 *Let $\mathcal{P} = \{\succ_i\}_{i \in \mathcal{I}}$ be an arbitrary ordinal preference profile. If $\{u_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (X1) and (X2), then for any $\delta > 0$, we have*

$$\lim_{I \rightarrow \infty} \text{Prob} \left(U_{\mathcal{I}}(a) \geq U_{\mathcal{I}}^* - \delta \text{ for all } a \in \text{Score}_{\mathbf{s}}(\mathcal{P}) \right) = 1. \quad (10)$$

Furthermore, if the fourth moment of λ is finite,²⁷ then there are constants $C_1, C_2 > 0$ (determined by λ and σ_c^2) such that, for any $p > 0$, if $I \geq C_1/p$ and $I \geq C_2/p\delta^2$, then $\text{Prob}[U_{\mathcal{I}}(a < U_{\mathcal{I}}^* - \delta) < p$, for all $a \in \text{Score}_{\mathbf{s}}(\mathcal{P})$.

Conditional Impartial Culture. Theorem 6 is in fact a consequence of Theorem 7; in effect, the Exogenous Preference model can be interpreted as the Endogenous Preference model, conditional on a *particular* realization of the (random) ordinal preference profile $\{\succ_i\}_{i \in \mathcal{I}}$. Theorem 7 says that limit (9) holds for any particular realization of $\{\succ_i\}_{i \in \mathcal{I}}$. Theorem 6 follows from this fact by integrating over all *possible* realizations of $\{\succ_i\}_{i \in \mathcal{I}}$. (See Appendix B for details.)

At the end of Section 4.1, we explained how the “Impartial Culture” model was a special case of the Endogenous Preferences model. If we condition on a particular realization of $\{\succ_i\}_{i \in \mathcal{I}}$, we obtain the *Conditional Impartial Culture* model. To be precise, let ρ be a probability measure on \mathbb{R} with finite variance and no atoms. For every voter i in \mathcal{I} , let \succ_i be voter i ’s (exogenous) ordinal preference relation on \mathcal{A} . In this case, hypothesis (X2) takes the following form:

²⁷The *fourth moment* of the multivariate probability measure λ is the integral $\int_{\mathbb{U}} \sum_{n=1}^N u_n^4 d\lambda[\mathbf{u}]$. It is finite if $d\lambda[\mathbf{u}]$ decays quickly enough as $\|\mathbf{u}\| \rightarrow \infty$. For example, the fourth moment of a multivariate normal probability measure is finite.

(X2') Let $\{r_1^i, r_2^i, \dots, r_N^i\}$ be a sample of N independent, ρ -random variables. Rearrange this sample in increasing order, to obtain $r_{(1)}^i < r_{(2)}^i < \dots < r_{(N)}^i$ (these are called the *order statistics* of the sample). If $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$ and $a_1 \prec_i a_2 \prec_i \dots \prec_i a_N$, then set $u_i(a_1) := r_{(1)}^i$, $u_i(a_2) := r_{(2)}^i$, \dots , and $u_i(a_N) := r_{(N)}^i$.

The rank scoring rule described prior to Theorem 7 now has the following construction. Take a random sample of N independent random variables drawn from ρ , and compute the order statistics of this sample; this yields N new random variables (which are neither independent, nor identically distributed). Let $s_1^N < s_2^N < \dots < s_N^N$ be the expected values of these random variables. Then set $\mathbf{s} := (s_1^N, s_2^N, \dots, s_N^N)$.

It is convenient to “renormalize” $s_1^N, s_2^N, \dots, s_N^N$ to range over the interval $[-1, 1]$, by defining

$$\tilde{s}_n^N := \frac{2s_n^N - s_N^N - s_1^N}{s_N^N - s_1^N}, \quad \text{for all } n \text{ in } [1 \dots N].$$

This ensures that $\tilde{s}_N^N = 1$ and $\tilde{s}_1^N = -1$. (For example, if $N = 3$, then we have $\tilde{s}_3^3 = 1$ and $\tilde{s}_1^3 = -1$, and only the value of \tilde{s}_2^3 remains to be determined.) If ρ is a probability distribution symmetrically distributed about some point in the real line, then the values $\tilde{s}_1^N, \tilde{s}_2^N, \dots, \tilde{s}_N^N$ will be symmetrically distributed around zero—that is, $\tilde{s}_k^N = -\tilde{s}_{N+1-k}^N$ for all k in $[1 \dots N]$. Thus, if N is odd and $k = (N + 1)/2$, then $\tilde{s}_k^N = 0$. In particular, if $N = 3$, then we must have $\tilde{s}_2^3 = 0$, while $\tilde{s}_3^3 = 1$ and $\tilde{s}_1^3 = -1$. Thus, we get the rank scoring rule defined by the scoring vector $(-1, 0, 1)$, which is just the Borda rule. Thus, Theorem 7 implies the next result, which says that the Borda rule is asymptotically utilitarian-optimal for *any* symmetric measure ρ .

Corollary 9 *Suppose $|\mathcal{A}| = 3$, and let $\mathcal{P} = \{\succ_i\}_{i \in \mathcal{I}}$ be any profile of preference orders on \mathcal{A} . Let ρ be any symmetric, finite-variance probability distribution on \mathbb{R} . If $\{u_i\}_{i \in \mathcal{I}}$, and $\{c_i\}_{i \in \mathcal{I}}$ satisfy (X1), (X2'), and (X3), then $\lim_{I \rightarrow \infty} \text{Prob} \left[\text{Borda}(\mathcal{P}) \subseteq \underset{\mathcal{A}}{\text{argmax}}(U_{\mathcal{I}}) \right] = 1$.*

If $|\mathcal{A}| \geq 4$, then the Borda rule is no longer guaranteed to be asymptotically optimal; the optimal rule will depend on the expected values of the order statistics for ρ , which depend on the structure of ρ itself. For example, suppose ρ was a normal probability distribution and $|\mathcal{A}| = 7$. Then we get the following expected order statistics (to 5 significant digits).²⁸

$$\begin{array}{ll} s_7^7 \approx 1.35218, & \tilde{s}_7^7 = 1, \\ s_6^7 \approx 0.75737, & \tilde{s}_6^7 \approx 0.56011, \\ s_5^7 \approx 0.35271, & \tilde{s}_5^7 \approx 0.26085, \\ s_4^7 = 0, & \text{which renormalize to } \tilde{s}_4^7 = 0, \\ s_3^7 \approx -0.35271, & \tilde{s}_3^7 \approx -0.26085, \\ s_2^7 \approx -0.75737, & \tilde{s}_2^7 \approx -0.56011, \\ \text{and } s_1^7 \approx -1.35218, & \text{and } \tilde{s}_1^7 = -1. \end{array}$$

²⁸Here we suppose for simplicity that ρ is a standard normal distribution. Any other normal distribution would yield the same values for $\tilde{s}_1^7, \dots, \tilde{s}_7^7$ after renormalization.

By comparison, the Borda rule uses the scoring vector $(-1, -0.6\bar{6}, -0.3\bar{3}, 0, 0.3\bar{3}, 0.6\bar{6}, 1)$.

Unfortunately, the expected values of order statistics are quite hard to compute for many probability distributions. Harter and Balakrishnan (1996) provide tables of these expected values for most of the common probability distributions (e.g. normal, exponential, Weibull, etc.); from this data it is easy to design the appropriate rank scoring rule.

5 Conclusion

This paper has neglected strategic voting and implementation issues. However, simulation results by Lehtinen (2007, 2008) suggest that strategic voting can often *improve* the utilitarian efficiency of some of the voting rules discussed in this paper. Furthermore, the results of Kim (2014) show that, with stochastically independent voters, the rank scoring rules considered in Section 4 are truth-revealing in Bayesian Nash equilibrium.²⁹ Finally, the central premise of our results is that it is sufficient to obtain an arbitrarily high *probability* of selecting a utilitarian optimum, rather than certainty. This is exactly the same premise as the theory of *virtual implementation* introduced by Matsushima (1988) and Abreu and Sen (1991). Virtual implementation is an extremely powerful and versatile implementation technology. For example, if the voters have complete information about one another, then *any* social choice rule can be virtually implemented in Nash equilibrium (Abreu and Sen, 1991) or iterated undominated strategies (Abreu and Matsushima, 1992). Even with incomplete information, a very large class of social choice rules can be virtually implemented in Bayesian Nash equilibrium (Serrano and Vohra, 2005), or even robustly virtually implemented (Artemov et al., 2013). Since virtual implementation is the implementation technology most suited to the probabilistic approach taken in this paper, we can consider the implementation problem to be essentially solved, for our purposes.

We have also assumed that the random variables associated with different voters are independent.³⁰ This excludes the possibility that voters belonging to a particular community or social group might exhibit correlations in their utility functions, preference intensities, errors, and/or approval thresholds. Empirical data suggests that these independence assumptions are false (Gelman et al., 2004). But they were made for expositional simplicity, rather than technical necessity. The results in this paper are derived using results from Section 2 of Pivato (2014b), which assumes stochastically independent voters. Section 3 of Pivato (2014b) extends these results to *correlated* voters, assuming the correlation strength is not too strong.³¹ Using this extension, it would be straightforward to extend the results of the present paper to correlated voters.

For a large enough population, our results suggest that simple and popular voting rules

²⁹Weber (1978) also states a special case of this result. However, a BNE with stochastically independent voters assumes that each voter is totally ignorant about the preferences of all other voters when she votes; this would not seem very realistic in many situations.

³⁰This assumption is shared by virtually all the literature reviewed in Section 1, except for Bordley (1985b, 1986), Fleurbaey (2009), and Beisbart and Hartmann (2010).

³¹To be more precise, we need an asymptotic condition on the L^1 norm of the covariance matrix of the random variables $\{c_i\}_{i \in \mathcal{I}}$ and $\{u_i\}_{i \in \mathcal{I}}$, as $I \rightarrow \infty$.

such as the Borda rule or approval voting have a very high probability of selecting an alternative which maximizes or almost-maximizes the utilitarian social welfare function. However, each result depends on specific assumptions about the statistical distribution of utility functions in the population. Thus, to select the best rule, we must know something about this distribution. The statistical distribution of utility functions probably depends on both the society and the particular policy problem. Thus, different voting rules will be optimal in different situations. In some situations, none of the voting rules considered here may be optimal, from a utilitarian perspective. This suggests a two-stage approach. In the first stage, estimate the utility functions of some statistically representative sample of the population (e.g. using a survey). Use this data to estimate the statistical distribution of utilities, and then determine which voting rule (if any) is optimal, given this distribution. If the statistical distribution of utility functions satisfies the hypotheses of one of the results in this paper, then in the second stage, we can use the corresponding voting rule to make the collective decision. Otherwise, we must resort to some other method —perhaps the methods considered in Pivato (2014a,b) or Kim (2014).

Acknowledgements. I am grateful to Michel le Breton, Rohan Dutta, Ori Heffetz, Sean Horan, Jérôme Lang, Michel le Breton, Christophe Muller, Matías Núñez, and Clemens Puppe for useful discussions and helpful comments on earlier versions of this paper. I also thank Gustaf Arrhenius, Miguel Ballester, Marc Fleurbaey, Annick Laruelle, and the other participants of the June 2014 “Workshop on Power” at the Collège d’Études Mondiales in Paris. None of these people are responsible for any errors. This research was supported by NSERC grant #262620-2008.

Appendix A: Background

The proofs in this paper depend on some results from Pivato (2014b). In this appendix, we briefly review these results.

Pivato (2014b) considers the problem of a utilitarian social planner who can only make noisy observations of the utility functions of the individuals in society and the correct system of interpersonal comparisons. For every i in \mathcal{I} , let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be the true cardinal utility function for voter i , and let $c_i > 0$ be a calibration constant, which we will use to make cardinal interpersonal utility comparisons. We suppose that the social planner wants to maximize the utilitarian social welfare function $U_{\mathcal{I}} : \mathcal{A} \rightarrow \mathbb{R}$ defined by formula (6); however, she faces the following informational problems.

- (U1) $\{c_i\}_{i \in \mathcal{I}}$ are unknown. The social planner regards $\{c_i\}_{i \in \mathcal{I}}$ as independent (but not necessarily identically distributed) real-valued random variables. There are constants $\bar{c} > 0$ and $\sigma_c^2 \geq 0$ such that $\mathbb{E}[c_i] = \bar{c}$ and $\text{var}[c_i] \leq \sigma_c^2$, for all $i \in \mathcal{I}$.
- (U2) The utility functions $\{u_i\}_{i \in \mathcal{I}}$ are not precisely observable. Instead, for each i in \mathcal{I} , the planner can only observe a function $v_i := u_i + \epsilon_i$, where $\epsilon_i : \mathcal{A} \rightarrow \mathbb{R}$ is a random “error” term. For each alternative a in \mathcal{A} , the random errors $\{\epsilon_i(a)\}_{i \in \mathcal{I}}$

are independent³² (but not necessarily identically distributed), and they all have an expected value of 0 and a variance less than or equal some constant $\sigma_\epsilon^2 > 0$.

Finally, the random variables $\{c_i\}_{i \in \mathcal{I}}$ are independent of the random functions $\{\epsilon_i\}_{i \in \mathcal{I}}$.

We assume that the utility profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies one or both of the following conditions.

(U3) There is a constant $\Delta > 0$ such that $\max_{\mathcal{A}}(U_{\mathcal{I}}) - U_{\mathcal{I}}(a) > \Delta$ for every $a \notin \operatorname{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$.

(U4) There is a constant $M > 0$ such that $\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 < M^2$ for every a in \mathcal{A} .

Note that, while we assume that $\{v_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ are random variables, we make no assumptions about the mechanism generating the underlying profile of utility functions $\{u_i\}_{i \in \mathcal{I}}$. These utility functions might be fixed in advance, or they might themselves be generated by some other random process, as long as they satisfy (U4) and (U3). Define the function $V_{\mathcal{I}} : \mathcal{A} \rightarrow \mathbb{R}$ as in equation (1), and define the SWF $U_{\mathcal{I}} : \mathcal{A} \rightarrow \mathbb{R}$ as in equation (6). Here is Theorem 1 of Pivato (2014b).

Theorem A1 *For every i in \mathcal{I} , let $u_i : \mathcal{A} \rightarrow \mathbb{R}$ be a utility function. Suppose the profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies (U3) and (U4). If $\{c_i\}_{i \in \mathcal{I}}$, $\{\epsilon_i\}_{i \in \mathcal{I}}$ and $\{v_i\}_{i \in \mathcal{I}}$ are randomly generated according to rules (U1) and (U2), then $\lim_{I \rightarrow \infty} \operatorname{Prob} \left[\operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{I}}) \subseteq \operatorname{argmax}_{\mathcal{A}}(U_{\mathcal{I}}) \right] = 1$.³³*

For any $\delta > 0$ and $p \in (0, 1)$, we define

$$\bar{I}(\delta, p) := 4A \frac{M^2 \sigma_c^2 + \sigma_\epsilon^2}{p \delta^2}. \quad (\text{A1})$$

Define $U_{\mathcal{I}}^* := \max \{U_{\mathcal{I}}(a) ; a \in \mathcal{A}\}$. Here is Theorem 2 of Pivato (2014b).

Theorem A2 *Suppose $\{u_i\}_{i \in \mathcal{I}}$, $\{c_i\}_{i \in \mathcal{I}}$, $\{\epsilon_i\}_{i \in \mathcal{I}}$ and $\{v_i\}_{i \in \mathcal{I}}$ satisfy (U1), (U2) and (U4). For any $\delta > 0$ and $p \in (0, 1)$, if $I \geq \bar{I}(\delta, p)$, then $\operatorname{Prob} [U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta] < p$, for all a in $\operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{I}})$.*

Appendix B: Proofs

Proof of Theorem 1. We will derive this from Theorem A1. Recall that $u_i := w_i/c_i$, so that $w_i = c_i u_i$; with this substitution, formulae (2) and (6) are equivalent. Observe that hypothesis (E1) implies (U1) (with $\bar{c} := 1$), hypothesis (E2) implies (U2), and hypothesis (E3) implies (U3). Meanwhile, hypothesis (U4) is true automatically, with $M = 1$, because the functions $\{u_i\}_{i \in \mathcal{I}}$ range over $[0, 1]$. The asymptotic probability claim now follows from Theorem A1. \square

³²We do *not* assume that, for a fixed voter i in \mathcal{I} , the random errors $\epsilon_i(a)$ and $\epsilon_i(b)$ are independent for different alternatives a and b in \mathcal{A} .

³³Here, the constants M and Δ in conditions (U3) and (U4) are to be held constant as $I \rightarrow \infty$.

Proof of Theorem 2. If we set $M := 1$ in formula (A1), we obtain formula (5). The asymptotic probability inequality now follows from Theorem A2. \square

Proof of Theorem 3. Let $g := \mathbb{E}[u_a^i | u_a^i \geq \theta_i]$ and let $b := \mathbb{E}[u_a^i | u_a^i < \theta_i]$; thus $g > b$. By $(\Theta 2)$ and $(\Theta 0)$, these values do not depend on i or a . For all $i \in \mathcal{I}$, define $v_i : \mathcal{A} \rightarrow \mathbb{R}$ by

$$v_i(a) := \begin{cases} g & \text{if } u_a^i \geq \theta_i; \\ b & \text{if } u_a^i < \theta_i. \end{cases} = \begin{cases} g & \text{if } a \in \mathcal{G}_i; \\ b & \text{otherwise.} \end{cases}$$

(The second equality is by equation (7).) Then define $\epsilon_i(a) := v_i(a) - u_i(a)$ for all $a \in \mathcal{A}$. Thus, $v_i = u_i + \epsilon_i$. By construction, $\mathbb{E}[u_i(a) | v_i(a)] = v_i(a)$, and thus $\mathbb{E}[\epsilon_i(a)] = 0$, for all $i \in \mathcal{I}$ and $a \in \mathcal{A}$. Let $\bar{c} := g - b$; then $\bar{c} > 0$. For all $i \in \mathcal{I}$, let $\tilde{u}_i := u_i/\bar{c}$ and $\tilde{c}_i := \bar{c} \cdot c_i$; thus, $c_i u_i = \tilde{c}_i \tilde{u}_i$. Thus, $U_{\mathcal{I}} = \frac{1}{I} \sum_{i \in \mathcal{I}} \tilde{c}_i \tilde{u}_i$. For all $i \in \mathcal{I}$, let $\tilde{v}_i := v_i/\bar{c}$ and $\tilde{\epsilon}_i := \epsilon_i/\bar{c}$; thus, $\tilde{v}_i = \tilde{u}_i + \tilde{\epsilon}_i$. Let $\tilde{V}_{\mathcal{I}} = \sum_{i \in \mathcal{I}} \tilde{v}_i$. Then $\tilde{V}_{\mathcal{I}} = V_{\mathcal{G}} + (\text{a constant})$. Thus, $\operatorname{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}}) = \operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{G}})$. But $\operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{G}}) = \operatorname{Appr}(\mathcal{G})$; thus, it suffices to compute the asymptotic probability that $\operatorname{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}}) \subseteq \operatorname{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$, using Theorem A1. To do this, we must verify hypotheses (U1)-(U4). First, let $\bar{u} := \mathbb{E}[u_a^i]$ and let $\sigma_u^2 := \operatorname{var}[u_a^i]$ for any $i \in \mathcal{I}$ and $a \in \mathcal{A}$. By hypothesis $(\Theta 2)$, these values are finite and independent of i and a . Let $M := \bar{u}^2 + \sigma_u^2$.

Claim 3.1: $\lim_{I \rightarrow \infty} \operatorname{Prob} \left(M \text{ and the profile } \{u_i\}_{i \in \mathcal{I}} \text{ satisfy condition (U4)} \right) = 1.$

Proof. Fix $a \in \mathcal{A}$. For all $i \in \mathcal{I}$, we have $\mathbb{E}[u_i^2(a)] = \bar{u}^2 + \sigma_u^2 = M^2$. Thus, $\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2$ is an average of I independent random variables (by $(\Theta 2)$), each with expected value M^2 . Thus, the Law of Large Numbers implies that

$$\lim_{I \rightarrow \infty} \operatorname{Prob} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 < M^2 \right] = 1.$$

Thus, since \mathcal{A} is finite, the claim follows. \diamond **Claim 3.1**

Hypotheses $(\Theta 0)$ and $(\Theta 2)$ imply that $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$ and $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ satisfy (U2). Hypothesis $(\Theta 1)$ implies that $\{\tilde{c}_i\}_{i \in \mathcal{I}}$ satisfies (U1), and hypothesis $(\Theta 3)$ implies that $\{\tilde{u}_i\}_{i \in \mathcal{I}}$ satisfies (U3) (with $\tilde{\Delta} := \Delta/\bar{c}$). Now Apply Theorem A1 to $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$, $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ and $\{\tilde{c}_i\}_{i \in \mathcal{I}}$ to derive the claimed asymptotic probability. \square

Proof of Theorem 4. The strategy is very similar to the proof of Theorem 3. Let g be the mean value of γ , and let b be the mean value of β ; thus $g > b$. For all $i \in \mathcal{I}$, define $v_i : \mathcal{A} \rightarrow \mathbb{R}$ by

$$v_i(a) := \begin{cases} g & \text{if } a \in \mathcal{G}_i; \\ b & \text{if } a \in \mathcal{B}_i. \end{cases}$$

Then define $\epsilon_i(a) := v_i(a) - u_i(a)$ for all $a \in \mathcal{A}$. Thus, $v_i = u_i + \epsilon_i$. By construction, $\mathbb{E}[u_i(a)] = v_i(a)$, and thus $\mathbb{E}[\epsilon_i(a)] = 0$, for all $i \in \mathcal{I}$ and $a \in \mathcal{A}$. Let $\bar{c} := g - b$; then $\bar{c} > 0$. For all $i \in \mathcal{I}$, let $\tilde{u}_i := u_i/\bar{c}$ and $\tilde{c}_i := \bar{c} \cdot c_i$; thus, $c_i u_i = \tilde{c}_i \tilde{u}_i$. Thus, $U_{\mathcal{I}} = \frac{1}{I} \sum_{i \in \mathcal{I}} \tilde{c}_i \tilde{u}_i$. For all $i \in \mathcal{I}$, let $\tilde{v}_i := v_i/\bar{c}$ and $\tilde{\epsilon}_i := \epsilon_i/\bar{c}$; thus, $\tilde{v}_i = \tilde{u}_i + \tilde{\epsilon}_i$. Let $\tilde{V}_{\mathcal{I}} = \sum_{i \in \mathcal{I}} \tilde{v}_i$. Then $\tilde{V}_{\mathcal{I}} = V_{\mathcal{G}} + (\text{a constant})$. Thus, $\operatorname{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}}) = \operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{G}})$. But $\operatorname{argmax}_{\mathcal{A}}(V_{\mathcal{G}}) = \operatorname{Appr}(\mathcal{G})$; thus, it suffices to compute the asymptotic probability that $\operatorname{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}}) \subseteq \operatorname{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$, using Theorem A1.

Let $M_g := g^2 + \operatorname{var}[\gamma]$ and $M_b := b^2 + \operatorname{var}[\beta]$. Let $M := \sqrt{\max\{M_g, M_b\}}$ and let $\tilde{M} := M/\bar{c}$.

Claim 4.1: $\lim_{I \rightarrow \infty} \operatorname{Prob} \left(\tilde{M} \text{ and the profile } \{\tilde{u}_i\}_{i \in \mathcal{I}} \text{ satisfy condition (U4)} \right) = 1$.

Proof. Fix $a \in \mathcal{A}$. For all $i \in \mathcal{I}$, if $a \in \mathcal{G}_i$, then $\mathbb{E}[u_i^2(a)] = M_g$. If $a \in \mathcal{B}_i$, then $\mathbb{E}[u_i^2(a)] = M_b$. Either way, $\mathbb{E}[u_i^2(a)] \leq M^2$. Thus, $\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2$ is an average of I independent random variables (by (S2)), each with expected value M^2 . Thus, the Law of Large Numbers implies that

$$\lim_{I \rightarrow \infty} \operatorname{Prob} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 < M^2 \right] = 1.$$

Since $\tilde{u}_i := u_i/\bar{c}$ for all $i \in \mathcal{I}$, it follows that

$$\lim_{I \rightarrow \infty} \operatorname{Prob} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} \tilde{u}_i(a)^2 < \tilde{M}^2 \right] = 1.$$

Thus, since \mathcal{A} is finite, the claim follows. ◇ Claim 4.1

Hypothesis (S1) implies that $\{\tilde{c}_i\}_{i \in \mathcal{I}}$ satisfies (U1). Define $\sigma_{\tilde{c}}^2 := \max\{\operatorname{var}(\gamma), \operatorname{var}(\beta)\}/\bar{c}^2$. Since $\tilde{\epsilon}_i = \tilde{u}_i - \tilde{v}_i$, it follows that $\operatorname{var}(\tilde{\epsilon}_i) \leq \sigma_{\tilde{c}}^2$ for all i . Thus, hypothesis (S2) implies that $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$ and $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ satisfy (U2). Finally, hypothesis (S3) implies that $\{\tilde{u}_i\}_{i \in \mathcal{I}}$ satisfy (U3) (with $\tilde{\Delta} := \Delta/\bar{c}$). Now apply Theorem A1 to $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$, $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ and $\{\tilde{c}_i\}_{i \in \mathcal{I}}$ to derive the claimed asymptotic probability. □

Proof of Proposition 5. Define $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$, $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ and $\tilde{V}_{\mathcal{I}}$ as in the proof of Theorem 4. From Theorem A2, along with Claim 4.1, we immediately obtain the limit equation (8). However, to obtain more precise estimates of the convergence speed, we must first estimate the speed of the convergence in Claim 4.1, using the next result.

Claim 5.1: *Suppose the fourth moments of γ and β are finite. Then there is some $C_1 > 0$ (determined by γ and β) such that, for any $p \in (0, 1)$, if $I > C_1/p$, then*

$$\operatorname{Prob} \left(\tilde{M} \text{ and } \{\tilde{u}_i\}_{i \in \mathcal{I}} \text{ violate condition (U4)} \right) < \frac{p}{2}.$$

The proof is very similar to the proof of Claim 8.1 in the proof of Proposition 8 (below).

Recall that $\{\tilde{u}_i\}_{i \in \mathcal{I}}$, $\{\tilde{v}_i\}_{i \in \mathcal{I}}$ and $\{\tilde{\epsilon}_i\}_{i \in \mathcal{I}}$ satisfy (U2), with $\sigma_\epsilon^2 := \max\{\text{var}(\gamma), \text{var}(\beta)\}/\bar{c}^2$. For any $\delta > 0$ and $p \in (0, 1)$, define $\bar{I}(\delta, p)$ as in equation (A1). Finally, define $C_2 := 8A(\widetilde{M}^2 \sigma_\epsilon^2 + \sigma_\epsilon^2)$. Thus, for any $p, \delta \in (0, 1)$, if $I > C_2/p\delta^2$, then $I > \bar{I}(\delta, p/2)$, so that, for any $a \in \text{Appr}(\mathcal{G}) = \text{argmax}_{\mathcal{A}}(\widetilde{V}_I)$, Theorem A2 says

$$\text{Prob} \left[U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta \mid \widetilde{M} \text{ and } \{\tilde{u}_i\}_{i \in \mathcal{I}} \text{ satisfy (U4)} \right] < \frac{p}{2}. \quad (\text{B1})$$

If $I > C_1/p$ also, then Claim 5.1 applies. This, together with inequality (B1), implies that $\text{Prob}[U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta] < \frac{p}{2} + \frac{p}{2} = p$, as desired. \square

Theorem 6 follows from Theorem 7, so we will prove that first.

Proof of Theorem 7. Since $\text{Score}_{\mathcal{s}}(\mathcal{P}) = \text{argmax}_{\mathcal{A}}(V_{\mathcal{P}}^{\mathcal{s}})$, it suffices to compute the asymptotic probability that $\text{argmax}_{\mathcal{A}}(V_{\mathcal{P}}^{\mathcal{s}}) \subseteq \text{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$, as $I \rightarrow \infty$. As usual, we will use Theorem A1. Hypothesis (X1) implies (U1). For all $i \in \mathcal{I}$ and all $a \in \mathcal{A}$, if we know that i ranks a in k th place (in particular, if we know the preference order \succ_i), then the expected value of $u_i(a)$, conditional on this information, is s_k . But, by definition, $v_i(a) = s_k$. Thus, $\mathbb{E}[u_i(a) \mid \succ_i] = v_i(a)$. Thus, if we define $\epsilon_i(a) := u_i(a) - v_i(a)$, then $\mathbb{E}[\epsilon_i(a) \mid \succ_i] = 0$. By hypothesis, the variance of the random variable $u_i(a)$ is finite; thus, the variance of $\epsilon_i(a)$ is finite. Finally, by hypothesis (X2), the random functions $\{u_i\}_{i \in \mathcal{I}}$ are independent of one another and independent of $\{c_i\}_{i \in \mathcal{I}}$. Thus, the random functions $\{\epsilon_i\}_{i \in \mathcal{I}}$ are independent of one another and independent of $\{c_i\}_{i \in \mathcal{I}}$. This establishes (U2). It remains to verify (U4).

Let $\widehat{\mathbf{u}} = (\widehat{u}_1, \dots, \widehat{u}_N) \in \mathbb{U}$ be a λ -random vector. The coordinates $\widehat{u}_1, \dots, \widehat{u}_N$ are themselves random variables (neither independent, nor identically distributed). Let $\sigma_1^2, \dots, \sigma_N^2$ denote their variances. Since λ has finite variance, it is easy to check that $\sigma_1^2, \dots, \sigma_N^2$ are all finite. Define $\sigma_\epsilon^2 := \max\{\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2\}$. Also, let $S := \max\{|s_1|, |s_2|, \dots, |s_N|\}$, and choose any $M > \sqrt{S^2 + \sigma_\epsilon^2}$.

Claim 7.1: $\lim_{I \rightarrow \infty} \text{Prob} \left(M \text{ and the profile } \{u_i\}_{i \in \mathcal{I}} \text{ satisfy condition (U4)} \right) = 1$.

Proof. Fix $a \in \mathcal{A}$. For all $i \in \mathcal{I}$, if a is ranked k th from the bottom by \succ_i , then $u_i(a)$ is a random variable with mean s_k and variance σ_k^2 . Thus,

$$\mathbb{E}[u_i^2(a)] = (s_k)^2 + \sigma_k^2 \leq S^2 + \sigma_\epsilon^2 < M^2. \quad (\text{B2})$$

Thus, for any $a \in \mathcal{A}$, the sum $\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2$ is an average of I independent random variables, each with expected value smaller than M^2 , by inequality (B2). Thus, regardless of how the preferences $\{\succ_i\}_{i \in \mathcal{I}}$ are obtained, the Law of Large Numbers implies that

$$\lim_{I \rightarrow \infty} \text{Prob} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 < M^2 \right] = 1.$$

Thus, since \mathcal{A} is finite, the claim follows. \diamond **Claim 7.1**

Finally, hypothesis (X3) implies that $\{u_i\}_{i \in \mathcal{I}}$ satisfy (U3). Now apply Theorem A1 to $\{u_i\}_{i \in \mathcal{I}}$, $\{v_i\}_{i \in \mathcal{I}}$, $\{\epsilon_i\}_{i \in \mathcal{I}}$ and $\{c_i\}_{i \in \mathcal{I}}$ to derive the limit (9). \square

Proof of Theorem 6. Define \mathbb{U} as in Section 4.2, and let λ be the conditionalization of μ on \mathbb{U} . In the Endogenous Preference model of Section 4.1, the ordinal preference profile $\mathcal{P} = \{\succ_i\}_{i \in \mathcal{I}}$ is a random variable (determined by the underlying cardinal utility profile $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$). However, if we fix a *particular* realization of \mathcal{P} , then *conditional* on this realization, the probability distribution of the cardinal profile $\{\mathbf{u}^i\}_{i \in \mathcal{I}}$ is described by λ via the Exogenous Preference model of Section 4.2. Thus, for any particular realization of \mathcal{P} , Theorem 7 implies that the limit (9) holds.³⁴ Thus, integrating over all possible realizations of \mathcal{P} , and applying Lebesgue’s Dominated Convergence Theorem, we conclude that the limit (9) holds unconditionally.³⁵ \square

Proof of Proposition 8. We will apply Theorem A2. Define $\{\epsilon_i\}_{i \in \mathcal{I}}$ and M as in the proof of Theorem 7. Claim 7.1 established that $\lim_{I \rightarrow \infty} \text{Prob}[M \text{ and the profile } \{u_i\}_{i \in \mathcal{I}} \text{ satisfy condition (U4)}] = 1$. However, to obtain the more precise estimate of convergence speed, we need the next observation.

Claim 8.1: *Suppose the fourth moment of λ is finite. Then there is some $C_1 > 0$ (determined by λ) such that, for any $p \in (0, 1)$, if $I > C_1/p$, then*

$$\text{Prob} \left(M \text{ and } \{u_i\}_{i \in \mathcal{I}} \text{ violate condition (U4)} \right) < \frac{p}{2}.$$

Proof. If the fourth moment of λ is finite, then there is some $C' > 0$ such that for any $a \in \mathcal{A}$, the fourth moments of each of the random variables $\{u_i(a)\}_{i \in \mathcal{I}}$ is less than C' . In other words, the second moments of each of the random variables $\{u_i(a)^2\}_{i \in \mathcal{I}}$ is less than C' . This implies that there is some $C'' > 0$ such that the variance of each of $\{u_i(a)^2\}_{i \in \mathcal{I}}$ is less than C'' . Also, these random variables are independent. Thus,

$$\text{var} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 \right] < \frac{C''}{I}. \quad (\text{B3})$$

Next, inequality (B2) says each of $\{u_i(a)^2\}_{i \in \mathcal{I}}$ has expected value less than M^2 . Thus,

$$\mathbb{E} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 \right] < M^2. \quad (\text{B4})$$

³⁴To be precise, we fix an *infinite* sequence $(\succ_n)_{n=1}^{\infty}$ of ordinal preferences. Then, for any particular value of I , we identify \mathcal{I} with $[1 \dots I]$ and let $\mathcal{P} = \{\succ_n\}_{n=1}^I$. Theorem 7 then applies for every possible infinite sequence.

³⁵The set of all possible infinite sequences $(\succ_n)_{n=1}^{\infty}$ has a natural sigma-algebra (generated by “cylinder sets”, which are defined by fixing values for any finite number of coordinates). The Endogenous Preference model defines a probability measure on this sigma algebra (in fact, it is a Bernoulli stochastic process). In the last step of the proof, we integrate with respect to this probability measure.

Thus, Chebyshev’s inequality and inequalities (B3) and (B4) imply that there is some $C_1 > 0$ (determined by C'') such that, for any $p > 0$, if $I > C_1/p$, then

$$\text{Prob} \left[\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 > M^2 \right] < \frac{p}{2|\mathcal{A}|}. \quad (\text{B5})$$

Now, if the profile $\{u_i\}_{i \in \mathcal{I}}$ and M to violate condition (U4), then $\frac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 > M^2$ for some $a \in \mathcal{A}$. Thus, adding together $|\mathcal{A}|$ copies of inequality (B5) proves the claim.

◇ **Claim 8.1**

For any $\delta > 0$ and $p > 0$, let $\bar{I}(\delta, p)$ be as in equation (A1). Finally, define $C_2 := 8A(M^2\sigma_c^2 + \sigma_\epsilon^2)$. Thus, for any $p, \delta \in (0, 1)$, if $I > C_2/p\delta^2$, then $I > \bar{I}(\delta, p/2)$, so that, for all $a \in \text{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}})$, Theorem A2 says that

$$\text{Prob} \left[U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta \mid M \text{ and } \{u_i\}_{i \in \mathcal{I}} \text{ satisfy (U4)} \right] < \frac{p}{2}. \quad (\text{B6})$$

If $I > C_1/p$ also, then Claim 8.1 applies. This, together with inequality (B6), implies that $\text{Prob}[U_{\mathcal{I}}(a) < U_{\mathcal{I}}^* - \delta] < \frac{p}{2} + \frac{p}{2} = p$ for any $a \in \text{argmax}_{\mathcal{A}}(\tilde{V}_{\mathcal{I}}) = \text{Score}_{\mathbf{s}}(\mathcal{P})$, as desired. □

References

- Abreu, D., Matsushima, H., 1992. Virtual implementation in iteratively undominated strategies: complete information. *Econometrica* 60 (5), 993–1008.
- Abreu, D., Sen, A., 1991. Virtual implementation in Nash equilibrium. *Econometrica* 59 (4), 997–1021.
- Alcantud, J. C. R., Laruelle, A., 2014. Disapproval voting: a characterization. *Soc. Choice Welf.* 43 (1), 1–10.
- Apesteguia, J., Ballester, M. A., Ferrer, R., 2011. On the justice of decision rules. *Rev. Econ. Stud.* 78 (1), 1–16.
- Artemov, G., Kunimoto, T., Serrano, R., 2013. Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine. *Journal of Economic Theory* 148 (2), 424 – 447.
- Azrieli, Y., Kim, S., 2014. Pareto efficiency and weighted majority rules. *International Economic Review* (to appear).
- Badger, W. W., 1972. Political individualism, positional preferences, and optimal decision-rules. In: Niemi and Weisberg (1972), pp. 34–59.
- Barberà, S., Jackson, M. O., 2006. On the weights of nations: assigning voting power to heterogeneous voters. *J Polit Econ* 114 (2), 317–339.

- Beisbart, C., Bovens, L., 2007. Welfarist evaluations of decision rules for boards of representatives. *Soc. Choice Welf.* 29 (4), 581–608.
- Beisbart, C., Bovens, L., Hartmann, S., 2005. A utilitarian assessment of alternative decision rules in the council of ministers. *European Union Politics* 6 (4), 395–418.
- Beisbart, C., Hartmann, S., 2010. Welfarist evaluations of decision rules under interstate utility dependencies. *Soc. Choice Welf.* 34 (2), 315–344.
- Bordley, R. F., 1983. A pragmatic method for evaluating election schemes through simulation. *American Political Science Review* 77, 123–141.
- Bordley, R. F., 1985a. A precise method for evaluating election schemes. *Public Choice* 46 (2), 113–123.
- Bordley, R. F., 1985b. Using factions to estimate preference intensity: Improving upon one person/one vote. *Public Choice* 45 (3), 257–268.
- Bordley, R. F., 1986. One person/one vote is not efficient given information on factions. *Theory and Decision* 21 (3), 231–250.
- Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A. D., Sheffet, O., 2012. Optimal social choice functions: A utilitarian view. In: *Proceedings of the 13th ACM Conference on Electronic Commerce. EC '12.* ACM, New York, NY, USA, pp. 197–214.
- Bovens, L. L., Hartmann, S., 2007. Welfare, voting and the constitution of a federal assembly. In: Galavotti, M. C., Scazzieri, R., Suppes, P. (Eds.), *Reasoning, rationality and probability.* CSLI Publications, Stanford.
- Brams, S. J., Fishburn, P. C., 1983. *Approval voting.* Birkhäuser Boston, Mass.
- Caragiannis, I., Procaccia, A. D., 2011. Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence* 175 (9–10), 1655 – 1671.
- Coughlin, P., 1992. *Probabilistic Voting Theory.* Cambridge Univ. Press, Cambridge.
- Curtis, R. B., 1972. Decision-rules and collective values in constitutional choice. In: Niemi and Weisberg (1972), pp. 23–33.
- Dhillon, A., 1998. Extended Pareto rules and relative utilitarianism. *Soc. Choice Welf.* 15, 521–542.
- Dhillon, A., Mertens, J.-F., 1999. Relative utilitarianism. *Econometrica* 67, 471–498.
- Dubey, P., Shapley, L. S., 1979. Mathematical properties of the Banzhaf power index. *Math. Oper. Res.* 4 (2), 99–131.
- Fleurbaey, M., 2009. One stake one vote, (preprint).
- Gelman, A., Katz, J. N., Bafumi, J., 2004. Standard voting power indexes do not work: An empirical analysis. *British J. Polit. Sci* 34, 657–74.

- Giles, A., Postl, P., August 2012. Equilibrium and welfare of two-parameter scoring rules. (preprint).
- Harter, H. L., Balakrishnan, N., 1996. CRC handbook of tables for the use of order statistics in estimation. CRC Press, Boca Raton, FL.
- Hillinger, C., 2005. The case for utilitarian voting. *Homo Oeconomicus* 23, 295–321.
- Kellett, J., Mott, K., Summer 1977. Presidential primaries: Measuring popular choice. *Polity* 9 (4), 528–537.
- Kim, S., 2014. Ordinal versus cardinal voting rules: a mechanism design approach. (preprint).
- Koriyama, Y., Laslier, J.-F., Macé, A., Treibich, R., 2013. Optimal apportionment. *Journal of Political Economy* 121 (3), 584–608.
- Laruelle, A., Valenciano, F., 2008. *Voting and Collective Decision-Making: Bargaining and Power*. Cambridge UP, Cambridge, UK.
- Laruelle, A., Valenciano, F., 2010. Egalitarianism and utilitarianism in committees of representatives. *Soc. Choice Welf.* 35 (2), 221–243.
- Ledyard, J. O., 1984. The pure theory of large two-candidate elections. *Public Choice* 44 (1), 7–41.
- Lehtinen, A., 2007. The Borda rule is also intended for dishonest men. *Public Choice* 133 (1-2), 73–90.
- Lehtinen, A., 2008. The welfare consequences of strategic behaviour under approval and plurality voting. *European Journal of Political Economy* 24, 688–704.
- Lindbeck, A., Weibull, J. W., 1987. Balanced-budget redistribution as the outcome of political competition. *Public Choice* 52 (3), 273–297.
- Lindbeck, A., Weibull, J. W., 1993. A model of political equilibrium in a representative democracy. *Journal of Public Economics* 51 (2), 195–209.
- Maaser, N., Napel, S., 2014. The mean voter, the median voter, and welfare-maximizing voting weights. In: R. Fara, D. L., Salles, M. (Eds.), *Voting Power and Procedures*. Studies in Choice and Welfare Series. Springer, Berlin.
- Macé, A., 2013. Voting with evaluations: When should we sum? What should we sum? (preprint).
- Macé, A., Treibich, R., 2012. Computing the optimal weights in a utilitarian model of apportionment. *Math. Soc. Sci* 63 (2), 141–51.
- Matsushima, H., 1988. A new approach to the implementation problem. *J. Econom. Theory* 45 (1), 128–144.
- McKelvey, R. D., Patty, J. W., 2006. A theory of voting in large elections. *Games Econom. Behav.* 57 (1), 155–180.

- Merrill, S., 1984. A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science* 28, 23–48.
- Myerson, R. B., 2002. Comparison of scoring rules in Poisson voting games. *J. Econom. Theory* 103 (1), 219–251.
- Niemi, R. G., Weisberg, H. F. (Eds.), 1972. *Probability Models of Collective Decision Making*. Charles E. Merrill Publishing, Columbus, Ohio.
- Nitzan, S., 2009. *Collective Preference and Choice*. Cambridge University Press.
- Núñez, M., Laslier, J.-F., 2014. Preference intensity representation: strategic overstating in large elections. *Social Choice and Welfare* 42 (2), 313–340.
- Ottowell, G., 1977. The arithmetic of voting. In *Defense of Variety* (July-August, 1977).
- Pivato, M., 2013. Voting rules as statistical estimators. *Social Choice and Welfare* 40 (2), 581–630.
- Pivato, M., 2014a. Condorcet meets Bentham. (preprint).
- Pivato, M., 2014b. Statistical utilitarianism. (preprint).
- Rae, D., 1969. Decision rules and individual values in constitutional choice. *Amer. Polit. Sci. Rev.* 63, 40–56.
- Riley, J., 1990. Utilitarian ethics and democratic government. *Ethics* 100 (2), 335–348.
- Schofield, N. J., 1972. Is majority rule special? In: Niemi and Weisberg (1972), pp. 60–82.
- Serrano, R., Vohra, R., 2005. A characterization of virtual Bayesian implementation. *Games Econom. Behav.* 50 (2), 312–331.
- Smith, W. D., 2000. Range voting, (unpublished; see <http://rangevoting.org>).
- Straffin, Jr., P. D., 1977. Majority rule and general decision rules. *Theory and Decision* 8 (4), 351–360.
- Taylor, M., 1969. Proof of a theorem on majority rule. *Behavioral Science* 14 (228-231).
- Weber, R. J., 1978. Comparison of voting systems. Discussion paper 498, Cowles Foundation for Research in Economics.