Thema
UMR 8184
THéorie Économique, Modélisation et Applications

Thema Working Paper n°2014-17
Université de Cergy Pontoise, France

"CONDORCET MEETS BENTHAM"

Marcus Pivato

September, 2014

UNIVERSITÉ de Cergy-Pontoise
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

CONDORCET MEETS BENTHAM

By Marcus Pivato[1]

Université de Cergy-Pontoise, France and Trent University, Canada

We show that if the statistical distribution of utility functions in a population satisfies a certain condition, then a Condorcet winner will not only exist, but will also maximize the utilitarian social welfare function. We also show that, if people's utility functions are generated according to certain plausible random processes, then in a large population, this condition will be satisfied with very high probability. Thus, in a large population, the utilitarian outcome will be selected by any Condorcet consistent voting rule. In particular, it will be the subgame-perfect equilibrium outcome of several voting games.

[1]Please address correspondence to: Marcus Pivato, THEMA, Université de Cergy-Pontoise, 33 Boulevard du Port, 95011 Cergy-Pontoise cedex, France. E-mail: *marcuspivato@gmail.com*.

1. Introduction

In the origins of modern social choice theory, one can distinguish two radically different approaches. One approach, due to Condorcet, focused on majority voting. The other, due to Bentham, focused on the maximization of social welfare, in the form of a utilitarian sum. The flaws of these two approaches are well-known. Utilitarianism requires complete knowledge of the utility functions of all individuals in society, and assumes a well-defined and unambiguous system of cardinal interpersonal utility comparisons. In the absence of this information, the utilitarian ideal is impossible to realize. But any procedure to acquire this information (e.g. via surveys) seems vulnerable to strategic manipulation. Finally, if the utilitarian choice is opposed by a large majority, then it may not be politically feasible.

Condorcet's approach does not suffer from these problems. Condorcet argued that society should choose a social alternative which is capable of beating any other single alternative in a majority vote. Such an alternative (if it exists) is called a *Condorcet winner*. A voting rule which always selects a Condorcet winner is called *Condorcet consistent*. Many well-known voting rules are Condorcet consistent, including the Copeland rule, the Simpson-Kramer rule, the Slater rule, the Kemeny rule, and any agenda of pairwise majority votes. Furthermore, if a Condorcet winner exists, then it will be the subgame-perfect Nash equilibrium outcome in *any* binary voting agenda (Miller, 1977, Proposition 8'), and many other multistage elimination procedures (Bag et al., 2009).[2] So Condorcet's approach is quite resistant to strategic voting. However, not all profiles of ordinal preferences admit a Condorcet winner. Furthermore, in general, there is no relationship between Condorcet consistency and social welfare.[3] So from a normative point of view, it is difficult to justify.

However, given a mild assumption (called "reasonability") about the statistical distribution of voter's preferences, we will show that the Condorcet winner actually maximizes utilitarian social welfare. We will then show that, if the voters' utility functions arise from certain plausible random processes, then a sufficiently large population of voters *will* have a reasonable distribution of utility functions, with very high probability. In other words, in a large population satisfying certain statistical regularities, not only is the Condorcet winner almost guaranteed to exist, but it is almost guaranteed to also be the utilitarian social choice. So for such populations, Condorcet and Bentham agree.

The remainder of this paper is organized as follows. Section 2 introduces basic notation and terminology, and states the foundational result: for "reasonable" utility profiles, the Condorcet winner is the utilitarian social choice. Section 3 considers a model where the utility functions of the voters are independent, identically distributed (i.i.d.) random variables drawn from a multivariate probability distribution with certain properties (e.g. a normal distribution). We show that, in a large population, the resulting profile of utility functions has a high probability of being reasonable. Section 4 considers spatial voting models, where the ideal points of the voters are i.i.d. random variables; again, under certain conditions, the resulting profile of utility functions has a high probability of being

[2] This assumes each voter has perfect information, and uses only weakly undominated strategies.

[3] Indeed, it is easy to construct examples where the Condorcet winner does *not* maximize social welfare (Lehtinen, 2007, §3).

reasonable for a large population. Finally, Section 5 reviews related literature. All proofs are in the Appendix.

2. Condorcet winners and reasonable utility profiles

Let $\mathcal{A}$ be a finite set of social alternatives, let $\mathcal{I}$ be a set of voters, and let $I := |\mathcal{I}|$. For every voter i in $\mathcal{I}$, let $u_i : \mathcal{A} \longrightarrow \mathbb{R}$ be i's cardinal utility function over $\mathcal{A}$. We refer to the set $\mathcal{U} := \{u_i\}_{i \in \mathcal{I}}$ as a *cardinal utility profile*. We will suppose that the utility functions $\{u_i\}_{i \in \mathcal{I}}$ admit one-for-one cardinal interpersonal comparisons.[4] Thus, a utilitarian would seek the social alternative which maximizes the utilitarian social welfare function $U_{\mathcal{I}}$ defined by

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

For every voter i in $\mathcal{I}$, let $\succ_i$ be the preference order induced by u_i on $\mathcal{A}$. We refer to the set $\mathcal{P} := \{\succ_i\}_{i \in \mathcal{I}}$ as an *ordinal preference profile*. Let $a \in \mathcal{A}$. We say that a is a *Condorcet winner* for $\mathcal{P}$ if, for every other alternative b in $\mathcal{A}$, some majority prefers a over b —that is, $\#\{i \in \mathcal{I};\ a \succ_i b\} \geq I/2$.

Let a and b be alternatives in $\mathcal{A}$, and for every voter i in $\mathcal{I}$, let $u^i_{a,b} := u_i(a) - u_i(b)$. Thus, $U_{\mathcal{I}}(a) \geq U_{\mathcal{I}}(b)$ if and only if the mean of the set $\mathcal{U}_{a,b} := \{u^i_{a,b}\}_{i \in \mathcal{I}}$ is positive. Meanwhile, a strict majority prefers a over b if and only if the *median* of $\mathcal{U}_{a,b}$ is positive. Thus, a strict majority will choose the $U_{\mathcal{I}}$-maximizing element of the pair $\{a, b\}$ if and only if $\mathrm{sign}[\mathrm{median}(\mathcal{U}_{a,b})] = \mathrm{sign}[\mathrm{mean}(\mathcal{U}_{a,b})]$. In this case, we say that the utility profile $\{u^i\}_{i \in \mathcal{I}}$ is *reasonable* relative to a and b.[5]

Example 1. If $|\mathrm{mean}(\mathcal{U}_{a,b})|$ exceeds the standard deviation of the set $\mathcal{U}_{a,b}$ (i.e. if the social welfare gap between the alternatives a and b is large enough), then the utility profile $\{u^i\}_{i \in \mathcal{I}}$ is (a, b)-reasonable. To see this, note that Chebyshev's inequality implies that $|\mathrm{median}(\mathcal{U}_{a,b}) - \mathrm{mean}(\mathcal{U}_{a,b})| \leq \mathrm{std\,dev}(\mathcal{U}_{a,b})$. ◇

We say the utility profile $\{u_i\}_{i \in \mathcal{I}}$ is *reasonable* if it is (a, b)-reasonable for every possible pair $a, b \in \mathcal{A}$. The following observation is immediate.

Theorem 2. *Let $\mathcal{U} = \{u_i\}_{i \in \mathcal{I}}$ be a cardinal utility profile, and let $\mathcal{P}$ be the corresponding ordinal preference profile. If $\mathcal{U}$ is reasonable, then $\mathcal{P}$ admits a Condorcet winner. This Condorcet winner maximizes the utilitarian social welfare function $U_{\mathcal{I}}$ in equation* (1).

[4] That is, for any alternatives $a, b, c, d \in \mathcal{A}$ and any voters $i, j \in \mathcal{I}$, if $u_i(b) - u_i(a) = u_j(d) - u_j(c)$, then the welfare that i gains in going from a to b is the same as the welfare that j gains in going from c to d.

[5] If I is odd, then $\mathrm{median}[\mathcal{U}_{ab}]$ is the unique point m in $\mathcal{U}_{a,b}$ such that $\#\{i \in \mathcal{I};\ u^i_{a,b} \geq m\} > I/2$ and $\#\{i \in \mathcal{I};\ u^i_{a,b} \leq m\} > I/2$. However, if I is *even*, then $\mathrm{median}[\mathcal{U}_{ab}]$ is generally an *interval* $[\underline{m}, \overline{m}]$ with $\underline{m} \leq \overline{m}$, such that $\#\{i \in \mathcal{I};\ u^i_{a,b} \geq \underline{m}\} \geq I/2$ and $\#\{i \in \mathcal{I};\ u^i_{a,b} \leq \overline{m}\} \geq I/2$. In this case, we will say $\mathrm{median}[\mathcal{U}_{ab}]$ is *positive* if $\overline{m} \geq \underline{m} > 0$, and we will say $\mathrm{median}[\mathcal{U}_{ab}]$ is *negative* if $\underline{m} \leq \overline{m} < 0$. If $\underline{m} \leq 0 \leq \overline{m}$, then we consider the "sign" of $\mathrm{median}[\mathcal{U}_{ab}]$ to be undefined (in this case, the voters are evenly split between a and b). Our definition of "reasonable" specifically excludes this last possibility.

Reasonability may seems like a heroic assumption, but the rest of this paper will show that it is actually quite plausible, under certain hypotheses. We will suppose that the voters' utility functions are randomly generated by some stochastic process. Under certain conditions, we shall see that, in a large population, such a randomly generated utility profile *will* be reasonable, with very high probability.

3. Random utility functions

Suppose $\mathcal{A}$ is a finite set, so that utility functions correspond to vectors in $\mathbb{R}^{\mathcal{A}}$. In this section, we will suppose that the voters' utility functions are i.i.d. random vectors. Here is an illustrative preliminary result.

Proposition 3. *Let ρ be any multivariate normal probability measure on $\mathbb{R}^{\mathcal{A}}$ with mean $\mathbf{m} \in \mathbb{R}^{\mathcal{A}}$ such that $m_a \neq m_b$ for any distinct a and b in $\mathcal{A}$. Suppose that the utility functions $\{u_i\}_{i\in\mathcal{I}}$ are independent, ρ-random variables. Then*

$$\lim_{I\to\infty} \text{Prob}\left(\textit{The utility profile } \{u_i\}_{i\in\mathcal{I}} \textit{ is reasonable}\right) = 1. \quad (2)$$

Thus, in a large enough population of voters with independent normally distributed utility functions, the Condorcet winner will exist, and will maximize utilitarian social welfare.

Proposition 3 raises two questions. First, how large must $\mathcal{I}$ be to ensure that the utility profile is reasonable with some probability (say, 95%)? Second, for what other probability distributions can we obtain a similar result? We will now answer these questions.

Let ρ be a probability measure on $\mathbb{R}^{\mathcal{A}}$. For any distinct alternatives a and b in $\mathcal{A}$, let $\rho_{a,b}$ be the distribution of the quantity $x_a - x_b$, where $\mathbf{x}$ is a ρ-random variable. We will say that the measure ρ is *reasonable* if ρ has finite variance, and if mean$[\rho_{a,b}]$ and median$[\rho_{a,b}]$ are nonzero and have the same sign, for all distinct alternatives a and b in $\mathcal{A}$. (For example, if ρ is any multivariate normal distribution satisfying the hypothesis in Proposition 3, then ρ is reasonable.) The next result generalizes Proposition 3; it says that reasonable measures generate reasonable utility profiles.

Theorem 4. *Let $\mathcal{A}$ be a finite set, let ρ be a reasonable probability measure on $\mathbb{R}^{\mathcal{A}}$, and suppose that the utility functions $\{u_i\}_{i\in\mathcal{I}}$ are independent, ρ-random variables. Then the limit* (2) *holds. To be precise, there are constants $q \in (0,1)$ and $C > 0$ (determined by the structure of ρ) such that, if I is large enough, then*

$$\text{Prob}\left(\{u_i\}_{i\in\mathcal{I}} \textit{ is } \text{not } \textit{reasonable}\right) < \frac{|\mathcal{A}|^2}{2}\left(2\sqrt{I}\,q^I + \frac{C}{I}\right) \xrightarrow[I\to\infty]{} 0. \quad (3)$$

REMARK 5. (a) Inequality (3) tells us how large I must be to ensure some probability that $\{u_i\}_{i\in\mathcal{I}}$ is reasonable. Note that $q^I \to 0$ very rapidly as $I\to\infty$. Thus, inequality (3) is dominated by the term $\frac{C\,|\mathcal{A}|^2}{2I}$. For example, suppose $q = 0.98$. If $I \geq 10\,000$, then $\sqrt{I}(0.98)^I \leq 10^{-85}$, so we can ignore it. Suppose $|\mathcal{A}| = 7$, $C = 10$, and $I = 10\,000$. Then

$$\text{Prob}\left(\{u_i\}_{i=1}^{10000} \text{ is not reasonable}\right) \;<\; \frac{49}{2}\left(2\sqrt{I}\,(0.98)^I + \frac{10}{I}\right) \;\approx\; \frac{25\cdot 10}{10\,000} \;=\; 0.025.$$

In other words, a ρ-random utility profile of ten thousand voters will be reasonable with probability at least 97.5%. Thus, with very high probability, the Condorcet winner of such a profile will be the utilitarian optimum.

(b) The condition that $m_a \neq m_b$ for all $a, b \in \mathcal{A}$ is not really essential in Proposition 3; it is for technical convenience. If $m_a = m_b$ for some $a, b \in \mathcal{A}$ *other* than the maximizer of $U_\mathcal{I}$, then the Condorcet winner will still maximize $U_\mathcal{I}$. If $m_a = m_b$ and one of them *is* the maximizer $U_\mathcal{I}$, then $U_\mathcal{I}(a)$ and $U_\mathcal{I}(b)$ will be very close, and *one* of them will be the Condorcet winner (with very high probability, when I is large). Thus, even if the Condorcet winner does not maximize $U_\mathcal{I}$, it will still "almost" maximize it. Similarly, the condition mean$[\rho_{a,b}] \neq 0$ is only for technical convenience in Theorem 4; even if ρ violates this condition, the Condorcet winner will either maximize or almost-maximize $U_\mathcal{I}$ (with very high probability, when I is large).

4. SPATIAL VOTING WITH RANDOM IDEAL POINTS

Spatial voting models are very common in the theoretical political science literature.[6] In these models, we regard $\mathbb{R}^N$ as a space of policies described by N distinct parameters. For example, different coordinates of $\mathbb{R}^N$ might represent interest rates, tax rates, expenditure levels for various public goods or income support mechanisms, and/or the inflation and unemployment rates. Suppose that each voter i in $\mathcal{I}$ has some "ideal point" $\mathbf{x}_i$ in $\mathbb{R}^N$. In this section, we will suppose that the voters' ideal points are i.i.d. random vectors, and that the utility that each voter assigns to a policy is a *decreasing function* of the distance from that policy to her ideal point. Thus, the voter prefers policies which are closer to her ideal point. Let $\mathbf{0} := (0, 0, \ldots, 0)$. Here is an illustrative preliminary result.

PROPOSITION 6. *Let ρ be any multivariate normal probability distribution on $\mathbb{R}^N$ with mean $\mathbf{0}$, and let $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ be independent ρ-random points. Let $\mathcal{A} \subset \mathbb{R}^N$ be a finite set of alternatives such that $\|\mathbf{a}\| \neq \|\mathbf{b}\|$ for any distinct $\mathbf{a}, \mathbf{b}$ in $\mathcal{A}$,[7] and suppose that $u_i(\mathbf{a}) = -\|\mathbf{a} - \mathbf{x}_i\|^2$ for every voter i in $\mathcal{I}$ and every alternative $\mathbf{a}$ in $\mathcal{A}$. Then the limit* (2) *holds.*

Thus, if $|\mathcal{I}|$ is sufficiently large, then Proposition 6 says that the utility profile $\{u_i\}_{i\in\mathcal{I}}$ will be reasonable, with very high probability. Thus, with very high probability, the Condorcet winner of such a profile will be the utilitarian optimum.

Proposition 6 has two limitations. First, it assumes a normal distribution of ideal points, and second, it assumes negative quadratic utility functions. In the rest of this

[6] See e.g. Hinich and Munger (1997) or Enelow and Hinich (2008) for introductions to this literature.
[7] Here, $\|\bullet\|$ is the Euclidean norm on $\mathbb{R}^N$.

section, we will work to relax these assumptions. We now let ρ be an arbitrary continuous probability measure on $\mathbb{R}^N$. We will use ρ to randomly generate the ideal points of the voters. We will suppose that each voter has a distance-based utility function of the form $u_i(\mathbf{a}) = -\phi(\|\mathbf{a} - \mathbf{x}_i\|)$ for some increasing function $\phi : [0, \infty) \longrightarrow \mathbb{R}$.

PROPOSITION 7. *Let $N \geq 2$. Let ρ be a continuous probability measure on $\mathbb{R}^N$ that is rotationally symmetric around $\mathbf{0}$, and let $\{\mathbf{x}_i\}_{i \in \mathcal{I}}$ be independent ρ-random points. Suppose $\phi : [0, \infty) \longrightarrow \mathbb{R}$ is strictly convex and increasing* (e.g. $\phi(x) = x^p$, for some $p > 1$), *and*

$$\int_{\mathbb{R}^N} \phi\left(\|\mathbf{y} - \mathbf{x}\|\right) \, \mathrm{d}\rho[\mathbf{x}] \quad < \quad \infty \quad \textit{for all } \mathbf{y} \textit{ in } \mathbb{R}^N. \tag{4}$$

Let $\mathcal{A} \subset \mathbb{R}^N$ be a finite set of alternatives such that $\|\mathbf{a}\| \neq \|\mathbf{b}\|$ for any distinct $\mathbf{a}, \mathbf{b}$ in $\mathcal{A}$. Suppose $u_i(\mathbf{a}) = -\phi(\|\mathbf{a} - \mathbf{x}_i\|)$ for every i in $\mathcal{I}$ and $\mathbf{a}$ in $\mathcal{A}$. Then the limit (2) *holds.*

If ρ is a standard normal distribution, then Proposition 6 is a consequence of Proposition 7. However, neither result is consequence of the other in general (because not all normal distributions are rotationally symmetric). In fact, Propositions 6 and 7 are both consequences of a single, more general result, as we now explain.

Let ρ be a continuous probability measure on $\mathbb{R}^N$. For any vector $\mathbf{v}$ in $\mathbb{R}^N$, a $\mathbf{v}$-*median hyperplane* of ρ is any hyperplane $\mathcal{H}^\rho_\mathbf{v} \subset \mathbb{R}^N$ which is orthogonal to $\mathbf{v}$, and such that at least half the mass of ρ lies on each side of $\mathcal{H}^\rho_\mathbf{v}$.[8] Such a hyperplane always exists,[9] but it might not be unique for some vectors $\mathbf{v}$ in $\mathbb{R}^N$. However, if there is a $\mathbf{v}$-median hyperplane $\mathcal{H}^\rho_\mathbf{v}$ which intersects the support of ρ, then $\mathcal{H}^\rho_\mathbf{v}$ is the *only* $\mathbf{v}$-median hyperplane.[10]

Let $\phi : [0, \infty) \longrightarrow \mathbb{R}$ be any convex increasing function. The ϕ-*median* of ρ is the set of global minima for the function $\Phi_\rho : \mathbb{R}^N \longrightarrow \mathbb{R}$ defined by

$$\Phi_\rho(\mathbf{m}) \quad := \quad \int_{\mathbb{R}^N} \phi\left(\|\mathbf{m} - \mathbf{x}\|\right) \, \mathrm{d}\rho[\mathbf{x}], \quad \text{for all } \mathbf{m} \text{ in } \mathbb{R}^N. \tag{5}$$

(For example, if $N = 1$ and $\phi(x) = x$ for all $x \geq 0$, then the ϕ-median of ρ is the classical median of ρ: the point(s) in $\mathbb{R}$ which cut the distribution of ρ into two equal halves.) We will say that ρ is ϕ-*balanced* if:

(B1) The function Φ_ρ is well-defined by the integral (5);[11]

(B2) The ϕ-median of ρ is a single point, $\mathbf{m}^\phi_\rho$;

(B3) Φ_ρ is rotationally symmetric around $\mathbf{m}^\phi_\rho$; and

---

[8] If $N = 1$, then the vector $\mathbf{v}$ is irrelevant, and a median "hyperplane" of ρ is actually a single point —it is any point h in $\mathbb{R}$ such that $\rho(-\infty, h] \geq \frac{1}{2}$ and $\rho[h, \infty) \geq \frac{1}{2}$.

[9] To see this, appy the Intermediate Value Theorem to the function f defined by $f(r) := \rho\{\mathbf{x} \in \mathbb{R}^N;\ \mathbf{v} \bullet \mathbf{x} \leq r\}$ (for all $r \in \mathbb{R}$), which is continuous because ρ is continuous.

[10] A point $\mathbf{x}$ in $\mathbb{R}^N$ is in the *support* of ρ if $\rho[\mathcal{U}] > 0$ for any open set $\mathcal{U} \subseteq \mathbb{R}^N$ which contains $\mathbf{x}$. Thus, $\mathcal{H}^\rho_\mathbf{v}$ intersects the support of ρ if and only if $\rho[\mathcal{U}] > 0$ for any open set $\mathcal{U} \subseteq \mathbb{R}^N$ which contains $\mathcal{H}^\rho_\mathbf{v}$.

[11] This is equivalent to inequality (4); it means that $\rho(\mathbf{x}) \to 0$ fast enough as $\|\mathbf{x}\| \to \infty$.

(B4) For every vector $\mathbf{v}$ in $\mathbb{R}^N$, there is a unique $\mathbf{v}$-median hyperplane $\mathcal{H}^{\mathbf{v}}_\rho$, and $\mathbf{m}^\phi_\rho \in \mathcal{H}^{\mathbf{v}}_\rho$.

For example, suppose $\phi(x) = x^2$ for all $x \geq 0$. If ρ has finite variance, then (B1) and (B2) are satisfied, and $\mathbf{m}^\phi_\rho$ is the mean of the distribution ρ. Indeed, a straightforward computation yields $\Phi_\rho(\mathbf{x}) = \mathrm{var}[\rho] + \left\|\mathbf{x} - \mathbf{m}^\phi_\rho\right\|^2$ for any $\mathbf{x}$ in $\mathbb{R}^N$.[12] Thus, condition (B3) is also satisfied. Thus, ρ is ϕ-balanced if and only if the mean of ρ lies in every median hyperplane of ρ. In particular:

- Any multivariate normal probability measure is ϕ-balanced. (See Lemma A2.)
- If ρ is a ϕ-balanced measure on $\mathbb{R}^N$, and $F : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ is any affine transformation, then $F(\rho)$ is a ϕ-balanced measure on $\mathbb{R}^M$. (*Proof:* F maps the mean of ρ to the mean of $F(\rho)$. Meanwhile, the F-preimage of any median hyperplane of $F(\rho)$ is a median hyperplane of ρ.)
- If $N = 1$, then ρ is ϕ-balanced if ρ has finite variance and is symmetrically distributed about some point m contained in the support of ρ. (For example, a uniform distribution on an interval is ϕ-balanced. So is the Laplace double-exponential distribution.)

Our last result says that, if any ϕ-balanced measure is used to generate a random collection of ideal points, which in turn is used to obtain a profile of distance-based utility functions, then this utility profile will be reasonable, with very high probability.

Theorem 8. *Let $\phi : [0, \infty) \longrightarrow \mathbb{R}$ be a convex increasing function, and let ρ be a ϕ-balanced probability measure on $\mathbb{R}^N$ with ϕ-median point $\mathbf{m}^\phi_\rho$. Let $\mathcal{A} \subset \mathbb{R}^N$ be a finite set of alternatives, such that $\left\|\mathbf{a} - \mathbf{m}^\phi_\rho\right\| \neq \left\|\mathbf{b} - \mathbf{m}^\phi_\rho\right\|$ for any distinct $\mathbf{a}, \mathbf{b}$ in $\mathcal{A}$. Finally, let $\{\mathbf{x}_i\}_{i \in \mathcal{I}}$ be a set of independent ρ-random points in $\mathbb{R}^N$. Suppose $u_i(\mathbf{a}) = -\phi\left(\|\mathbf{a} - \mathbf{x}_i\|\right)$ for every voter i in $\mathcal{I}$ and every $\mathbf{a}$ in $\mathcal{A}$. Then the limit* (2) *holds.*

Remark. (a) The condition "$\|\mathbf{a}\| \neq \|\mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$" is not really necessary in Propositions 6 and 7; it is for technical convenience. The same is true for requirement in Theorem 8 that $\left\|\mathbf{a} - \mathbf{m}^\phi_\rho\right\| \neq \left\|\mathbf{b} - \mathbf{m}^\phi_\rho\right\|$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$. If these conditions are violated for some $\mathbf{a}, \mathbf{b} \in \mathcal{A}$ *other* than the maximizer of $U_\mathcal{I}$, then the Condorcet winner will still maximize $U_\mathcal{I}$. If one of $\mathbf{a}$ or $\mathbf{b}$ *is* the maximizer of $U_\mathcal{I}$, then $U_\mathcal{I}(\mathbf{a})$ and $U_\mathcal{I}(\mathbf{b})$ will be very close, and *one* of them will be the Condorcet winner (with very high probability, when I is large). Thus, the Condorcet winner will either maximize $U_\mathcal{I}$, or almost-maximize it.

(b) In a general spatial voting model, McKelvey et al. (1980, Thm.2) give a necessary and sufficient condition for the existence of a Condorcet winner, which is similar to condition (B4).[13] The difference is that they apply this condition to the *actual* distribution of ideal points, whereas we apply it to the underlying probability distribution from which these ideal points are drawn. In their model, the Condorcet winner *is* the median point $\mathbf{m}^\phi_\rho$, whereas in our model, (B4) implies that the Condorcet winner is the alternative in

[12] This result is sometimes attributed to Christiaan Huygens.

[13] I thank Michel le Breton for pointing out this connection.

$\mathcal{A}$ which is closest to $\mathbf{m}^\phi_\rho$, while (B1)-(B3) imply that this same alternative maximizes the utilitarian SWF (with high probability, as $I\to\infty$).

5. Related literature

The results in this paper complement 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 focused on Condorcet consistent rules, Pivato (2014a) focuses on scoring rules such as the Borda rule or approval voting. 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.

The results in this paper are also reminiscent of the Condorcet Jury Theorem (CJT), and the literature it has generated.[14] Like the CJT, this paper says that, under certain statistical assumptions, a large population using a certain voting rule is likely to make the "correct" decision. But the goal of the CJT is to find the correct answer to an objective factual question, whereas the goal in the present paper is to maximize social welfare.

The utilitarian analysis of majority voting was pioneered by Rae (1969) and Taylor (1969). Assuming voters had i.i.d. $\{0,1\}$-valued utility functions over two alternatives, they showed that, amongst all anonymous voting rules, simple majority vote maximized the expected value of the utilitarian SWF. This result has been extended to simple games and weighted majority rules by Badger (1972), Curtis (1972), Schofield (1972), Straffin (1977), Dubey and Shapley (1979), Bordley (1985, 1986), Fleurbaey (2009), and Laruelle and Valenciano (2010).

More recently, Schmitz and Tröger (2012) have shown that "weak" majority voting rules yield the highest expected value for the utilitarian SWF amongst all dominant-strategy rules. As in the present paper, Schmitz and Tröger (2012) assume all voters are *ex ante* identical in the distribution of their utility functions. Azrieli and Kim (2014) relax this assumption, so that different voters may have different preference intensities, *ex ante.* Assuming 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).

The aforementioned papers considered only dichotomous decisions. But Lehtinen (2007, 2014) has considered agendas of pairwise votes involving any number of alternatives; using computer simulations, he showed that strategic voting generally *improves* the utilitarian social welfare of the outcome, in settings with incomplete information. Most recently, assuming voters with independent (but not identically distributed) random utility functions

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

over any number of alternatives, Kim (2014) characterized the rules which are *ex ante* Pareto efficient in the class of ordinal voting rules: they are "non-anonymous" scoring rules (where each voter has perhaps a different score vector). He further showed that, if the alternatives are *ex ante* interchangeable, then such rules are truth-revealing in Bayesian Nash equilibrium (BNE). A special case are the scoring rules which maximize expected utilitarian social welfare over all ordinal rules. Kim also constructed a rule which obtains a higher expected utilitarian social welfare than any ordinal rule in BNE.

Kim's rules do not always choose the utilitarian-optimal alternative —they just yield the highest *expected* utilitarian social welfare amongst all BNE-truth-revealing rules. In contrast, our Theorem 2 says that any Condorcet consistent rule will choose the utilitarian-optimal alternative in *any* reasonable profile. Theorem 4 shows that this is highly likely in a large population of voters with i.i.d. random utility functions —a model very similar to Kim's. One difference is that Kim's voters are not necessarily identically distributed *ex ante*, but they all have the same preference intensity *ex post* (i.e. every utility function ranges from 0 to 1). In contrast, our voters are i.i.d. *ex ante*, but may have different preference intensities *ex post*. Another difference is that the hypotheses of Theorem 4 contain a built-in asymmetry between the alternatives, whereas in Kim's model they are *ex ante* interchangeable.[15]

In effect, Theorem 2 yields an implementation of utilitarianism with informational assumptions diametrically opposite to Kim's. In an environment with independent random voters, a BNE means that each person votes in complete ignorance of the preferences of everyone else. In contrast, the Condorcet winner (when it exists) will be the subgame-perfect Nash equilibrium outcome of any agenda of pairwise votes (and several other "successive elimination" rules) when voters have *perfect* information about each other's preferences (Miller, 1977; Bag et al., 2009).[16] Thus, Theorems 2 and 4 together imply that, in a large population, with a distribution of utility functions similar to Kim (2014), any of these voting rules will provide a subgame-perfect implementation of utilitarianism.

6. Conclusion

This paper shows that, if the statistical distribution of utility functions in a large society satisfies certain conditions, then, with very high probability, a Condorcet winner will exist, and will maximize the utilitarian SWF. But in reality, does the distribution of utility functions in a particular society satisfy these conditions? This is an empirical question, and the answer probably depends on both the society and the particular policy problem under consideration. This suggests a two-stage approach to utilitarian social choice. In the first stage, use a survey or some other method to estimate the utility functions of a statistically representative sample of the population (measured, e.g. in terms of willingness-to-pay). Using this survey data, we can determine whether the distribution of utility functions is, in

[15]But as noted in Remark 5(b), this asymmetry is not essential to obtain close-to-optimal social welfare.

[16]Actually, a voter does not need *perfect* information; she just needs enough to reliably predict the outcome of each pairwise vote, so that she can correctly perform backwards induction. It is enough to have statistics about other voters' preferences, which could be obtained from public opinion polls.

fact, reasonable. If it *is* reasonable, then in the second stage, we can deploy any strategy-proof, Condorcet consistent social choice rule (e.g. an agenda of pairwise votes) to find the alternative which maximizes the utilitarian SWF. Otherwise, we must resort to some other method —e.g. the methods explored in Pivato (2014a,b) or Kim (2014).

In the models of Sections 3 and 4, one possibly questionable assumption is that the cardinal utility profile is a set of *independent* random variables.[17] This neglects the fact that voters belonging to the same community or subculture may exhibit correlations in their preferences. Empirical evidence suggests that the independence hypothesis is false (Gelman et al., 2004). However, full independence is not required for our results. The stochastic process generating the utility profile *can* have correlations, as long as the sample mean and sample medians converge to the mean and medians of the underlying distribution as $I \to \infty$. For example, this is true for any ergodic stochastic process. It will also happen if the correlations between voters are sufficiently weak; see Pivato (2014b) for an illustration of this approach.

If a utility profile $\mathcal{U}$ is *not* reasonable, then Theorem 2 does not apply; there may be no Condorcet winner, and even if there is, the Condorcet winner is not guaranteed to be a utilitarian optimum. However, if $\mathcal{U}$ is "close" to reasonable, then a suitably chosen Condorcet-consistent voting rule may still have a high probability of selecting a utilitarian optimum. For example, consider the *Copeland rule*, which chooses the alternative with the highest Copeland score. (The *Copeland score* of an alternative a is defined as $\#\{b \in \mathcal{A};$ some majority prefers a over $b\} - \#\{b \in \mathcal{A};$ some majority prefers b over $a\}$.) Suppose that, for every $a, b \in \mathcal{A}$, there is a small probability that the profile $\mathcal{U}$ will fail to be (a, b)-reasonable, and that this probability is decreasing as a function of the average utility gap between a and b (as suggested by Example 1). Also suppose that these reasonability failures are independent random variables. Then the Copeland score of each alternative should be a good estimator of the "true" ranking of that alternative by the utilitarian social welfare order. Thus, the Copeland winner should either be optimal or close-to-optimal with respect to the utilitarian social welfare order. By a similar argument, the ordering of $\mathcal{A}$ determined by the Slater rule should be a good estimate of the ordering of $\mathcal{A}$ determined by the utilitarian social welfare order. These are interesting questions for future research.

Acknowledgements. I am grateful to Michel le Breton, Sean Horan, 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

Proposition 3 is a special case of Theorem 4, so it suffices to prove the latter result.

[17] This assumption is shared by virtually all the literature reviewed in Section 5, except for Bordley (1985, 1986) and Fleurbaey (2009).

Proof of Theorem 4. For any distinct $a, b \in \mathcal{A}$, recall that $\rho_{a,b}$ is the distribution of $x_a - x_b$, where $\mathbf{x}$ is a ρ-random variable. Thus, $\rho_{a,b}$ has finite variance, because ρ has finite variance. Let $m_{a,b}$ be the mean value of $\rho_{a,b}$; then $m_{a,b} \neq 0$, because ρ is reasonable. Let $p_{a,b} := \rho_{a,b}(-\infty, 0)$ if $m_{a,b} > 0$, and let $p_{a,b} := \rho_{a,b}(0, \infty)$ if $m_{a,b} < 0$. (Equivalently, $p_{a,b} := \rho\{\mathbf{x} \in \mathbb{R}^{\mathcal{A}};\ \mathrm{sign}(x_a - x_b) = -\mathrm{sign}(m_{ab})\}$.) Then $p_{a,b} < \frac{1}{2}$, because $\mathrm{sign}(m_{a,b}) = \mathrm{sign}(\mathrm{median}[\rho_{a,b}])$, because ρ is reasonable. Let $p := \max\{p_{a,b};\ a, b \in \mathcal{A}\}$; then $p < \frac{1}{2}$ because $p_{a,b} < \frac{1}{2}$ for all $a, b \in \mathcal{A}$, and $\mathcal{A}$ is finite. It follows that $p(1-p) < \frac{1}{4}$ (because the function $f(x) = x(1-x)$ has a unique maximum at $x = \frac{1}{2}$, and $f(\frac{1}{2}) = \frac{1}{4}$). Thus, if we define $q := 2\sqrt{p(1-p)}$, then $q < 1$. (For example, if $p = 0.4$, then $q = 2\sqrt{0.4 \cdot 0.6} \approx 0.98$.) Let $A := |\mathcal{A}|$, and without loss of generality, suppose $\mathcal{A} = \{1, 2, \ldots, A\}$. For all $i \in \mathcal{I}$, let $u_i := (u_1^i, u_2^i, \ldots, u_A^i) \in \mathbb{R}^A$ be the utility function of voter i (a ρ-random vector). For any $a < b \in \mathcal{A}$, let $\mathcal{U}_{a,b} := \{u_a^i - u_b^i\}_{i \in \mathcal{I}}$ (a collection of I independent real-valued random variables).

Claim 1: *If I is large enough, then for all distinct $a, b \in \mathcal{A}$, we have*

$$\mathrm{Prob}\left[\mathrm{sign}\left(\mathrm{median}[\mathcal{U}_{a,b}]\right) \neq \mathrm{sign}\left(\mathrm{median}[\rho_{a,b}]\right)\right] \quad < \quad 2\sqrt{I}\, q^I.$$

Proof. Without loss of generality, suppose $\mathrm{median}(\rho_{a,b}) > 0$. Let J be the smallest integer greater than $I/2$. (That is: $J := (I+1)/2$ if I is odd, whereas $J := (I/2) + 1$ if I is even.) Now, $|\mathcal{U}_{a,b}| = I$ and $\mathrm{median}(\rho_{a,b}) > 0$, so[18]

$$\Big(\mathrm{sign}\left(\mathrm{median}[\mathcal{U}_{a,b}]\right) \neq \mathrm{sign}\left(\mathrm{median}[\rho_{a,b}]\right)\Big) \iff \Big(\mathrm{median}(\mathcal{U}_{a,b}) < 0\Big)$$
$$\iff \Big(\text{at least } J \text{ elements of } \mathcal{U}_{a,b} \text{ are in } (-\infty, 0)\Big). \qquad \text{(A1)}$$

Thus, we need to estimate the probability of the right hand side of (A1).

Let $\mathcal{U}' := \{x_1, x_2, \ldots, x_{2J}\}$ be a set of $2J$ i.i.d. random variables with $\mathrm{Prob}[x_k < 0] = p$ and $\mathrm{Prob}[x_k \geq 0] = 1 - p$, for all $k \in [1 \ldots 2J]$. Thus, for any $i \in \mathcal{I}$ and $k \in [1 \ldots 2J]$, we have $\mathrm{Prob}[u_a^i - u_b^i < 0] = p_{a,b} \leq p = \mathrm{Prob}[x_k < 0]$, and these are independent random events. Furthermore, $|\mathcal{U}'| > |\mathcal{U}_{a,b}|$ (because $2J > I$). Thus,

$$\mathrm{Prob}\Big(\text{at least } J \text{ elements of } \mathcal{U}_{a,b} \text{ are in } (-\infty, 0)\Big)$$
$$\leq \mathrm{Prob}\Big(\text{at least } J \text{ elements of } \mathcal{U}' \text{ are in } (-\infty, 0)\Big), \qquad \text{(A2)}$$

so it suffices to estimate the right hand side of inequality (A2). Now, for any $n \in$

[18]See footnote 5 for how to interpret the left-hand side of statement (A1) when I is even.

$[0 \ldots 2J]$,

$$\text{Prob}\Big(\text{exactly } n \text{ elements of } \mathcal{U}' \text{ are in } (-\infty, 0)\Big) \;=\; \binom{2J}{n} p^n (1-p)^{2J-n}. \quad \text{Thus,}$$

$$\begin{aligned}
&\text{Prob}\Big(\text{at least } J \text{ elements of } \mathcal{U}' \text{ are in } (-\infty, 0)\Big) \\
&\leq \sum_{n=J}^{2J} \binom{2J}{n} p^n (1-p)^{2J-n} \;\underset{\text{(a)}}{\leq}\; \sum_{n=J}^{2J} \binom{2J}{J} p^J (1-p)^J \\
&= J \binom{2J}{J} p^J (1-p)^J \;\underset{\text{(b)}}{<}\; J \frac{(2J)!}{(J!)^2} \Big(p(1-p)\Big)^{I/2} \\
&\underset{\text{(c)}}{=} J \frac{(2J)!}{(J!)^2} \left(\frac{q}{2}\right)^I \;\underset{(d)}{\approx}\; J \sqrt{\frac{2}{\pi I}} \cdot 2^{I+2} \left(\frac{q}{2}\right)^I \\
&= 4J \sqrt{\frac{2}{\pi I}} \cdot q^I \;\underset{\text{(e)}}{<}\; 2I \frac{q^I}{\sqrt{I}} \;=\; 2\sqrt{I}\, q^I.
\end{aligned} \tag{A3}$$

Here, (a) is because $p < \frac{1}{2}$, so the mode of the p-binomial distribution on $[0 \ldots 2J]$ occurs at some $n < J$, so that $\binom{2J}{n} p^n (1-p)^{2J-n} < \binom{2J}{J} p^J (1-p)^J$ for all $n \in [J \ldots 2J]$. Next, (b) is because $J > I/2$, and (c) is because $\sqrt{p(1-p)} = q/2$, so $[p(1-p)]^{I/2} = (\sqrt{p(1-p)})^I = (q/2)^I$. Next, (d) is via Stirling's approximation of the factorial, which says $n! \approx \sqrt{2\pi n}\,(n/e)^n$ as $n \to \infty$. Thus, if J is large enough, then

$$\frac{(2J)!}{(J!)^2} \;\approx\; \frac{\sqrt{2\pi\, 2J}\,(2J/e)^{2J}}{[\sqrt{2\pi\, J}\,(J/e)^J]^2} \;=\; \frac{2^{2J}}{\sqrt{\pi J}} \;<\; \frac{2^{I+2}}{\sqrt{\pi I/2}} \;=\; \sqrt{\frac{2}{\pi I}} \cdot 2^{I+2}.$$

Finally, (e) is because $2J \leq I + 2$ and $2 \cdot \sqrt{2/\pi} \approx 1.59$, so $4J \cdot \sqrt{2/\pi} \approx 2J\,(1.59) \leq (1.59)\,(I+2) < 2I$, if I is large enough.

Combining statement (A1) and inequalities (A2) and (A3) yields the claim. ◇ `Claim 1`

Let $C := \max\{\frac{\text{var}[\rho_{a,b}]}{m_{a,b}^2};\; a, b \in \mathcal{A}\}$; then $C < \infty$ because $\text{var}[\rho_{a,b}] < \infty$ and $m_{a,b} \neq 0$ for all distinct $a, b \in \mathcal{A}$, and $|\mathcal{A}|$ is finite.

Claim 2: *For all $a, b \in \mathcal{A}$,* $\text{Prob}\Big[\text{sign}\,(\text{mean}[\mathcal{U}_{a,b}]) \neq \text{sign}(m_{a,b})\Big] \leq C/I$.

Proof. Let $M_{a,b} := \text{mean}[\mathcal{U}_{a,b}] = \frac{1}{I}\sum_{i \in \mathcal{I}}(u_a^i - u_b^i)$. This is an avarage of i.i.d. random variables, each with expected value $m_{a,b}$ and variance $\text{var}[\rho_{a,b}]$. Thus, $M_{a,b}$ is itself a random variable with expected value $m_{a,b}$ and variance $\text{var}[\rho_{a,b}]/I$. Thus,

$$\text{Prob}\Big[\text{sign}\,(M_{a,b}) \neq \text{sign}(m_{a,b})\Big] \;\leq\; \text{Prob}\Big[|M_{a,b} - m_{a,b}| > m_{a,b}\Big] \;\underset{(*)}{\leq}\; \frac{\text{var}[\rho_{a,b}]}{I\, m_{a,b}^2} \;\leq\; \frac{C}{I},$$

as claimed. Here, $(*)$ is by Chebyshev's inequality. ◇ `Claim 2`

Now, sign (median$[\rho_{a,b}]$) = sign$(m_{a,b})$, because ρ is reasonable. Thus, if sign (median$[\mathcal{U}_{a,b}]$) = sign (median$[\rho_{a,b}]$) and sign (mean$[\mathcal{U}_{a,b}]$) = sign$(m_{a,b})$, then sign (median$[\mathcal{U}_{a,b}]$) = sign (mean$[\mathcal{U}_{a,b}]$). Conversely, if sign (median$[\mathcal{U}_{a,b}]$) $\neq$ sign (mean$[\mathcal{U}_{a,b}]$), then either sign (median$[\mathcal{U}_{a,b}]$) $\neq$ sign (median$[\rho_{a,b}]$) or sign (mean$[\mathcal{U}_{a,b}]$) $\neq$ sign$(m_{a,b})$. Thus,

$$\begin{aligned}
&\text{Prob}\Big[\text{sign}\,(\text{median}[\mathcal{U}_{a,b}]) \neq \text{sign}\,(\text{mean}[\mathcal{U}_{a,b}])\Big] \\
&\quad\leq\ \text{Prob}\begin{pmatrix}\text{sign}\,(\text{median}[\mathcal{U}_{a,b}]) \neq \text{sign}\,(\text{median}[\rho_{a,b}]) \\ \text{or}\quad \text{sign}\,(\text{mean}[\mathcal{U}_{a,b}]) \neq \text{sign}(m_{a,b})\end{pmatrix} \\
&\quad\leq\ \text{Prob}\Big[\text{sign}\,(\text{median}[\mathcal{U}_{a,b}]) \neq \text{sign}\,(\text{median}[\rho_{a,b}])\Big] \\
&\qquad\qquad + \ \text{Prob}\Big[\text{sign}\,(\text{mean}[\mathcal{U}_{a,b}]) \neq \text{sign}(m_{a,b})\Big] \\
&\quad\underset{(*)}{<}\ 2\sqrt{I}\,q^I + \frac{C}{I}, \qquad\qquad (\text{A4})
\end{aligned}$$

where $(*)$ is by Claims 1 and 2. Thus,

$$\begin{aligned}
&\text{Prob}\Big(\text{the profile } \{u_i\}_{i\in\mathcal{I}} \text{ is not reasonable}\Big) \\
&\quad=\ \text{Prob}\Big(\text{sign}\,(\text{median}[\mathcal{U}_{a,b}]) \neq \text{sign}\,(\text{mean}[\mathcal{U}_{a,b}]) \text{ for some } a<b\in\mathcal{A}\Big) \\
&\quad\underset{(*)}{<}\ \sum_{a<b\in\mathcal{A}} \left(2\sqrt{I}\,q^I + \frac{C}{I}\right) \ =\ \frac{A(A-1)}{2}\left(2\sqrt{I}\,q^I + \frac{C}{I}\right) \ \xrightarrow[I\to\infty]{(\dagger)}\ 0,
\end{aligned}$$

as desired. Here, the inequality $(*)$ follows from inequality (A4), and the limit $(\dagger)$ is a straightforward application of l'Hospital's rule, because $0 < q < 1$. □

Propositions 6 and 7 are special cases of Theorem 8, so we will prove that first. The proof of Theorem 8 and Proposition 7, in turn, use the following lemma.

Lemma A1. *Let ρ be any probability measure on $\mathbb{R}^N$, let $\phi : [0,\infty)\longrightarrow\mathbb{R}$ be any (strictly) convex function, and let Φ_ρ be defined as in equation* (5). *Then Φ_ρ is (strictly) convex.*

Proof. For any $\mathbf{x} \in \mathbb{R}^N$, define $\phi_\mathbf{x} : \mathbb{R}^N\longrightarrow\mathbb{R}$ by setting $\phi_\mathbf{x}(\mathbf{y}) := \phi\,(\|\mathbf{x}-\mathbf{y}\|)$ for all $\mathbf{y} \in \mathbb{R}^N$. First observe that $\phi_\mathbf{x}$ is (strictly) convex. To see this, let $\mathbf{y}, \mathbf{z} \in \mathbb{R}^N$, and let $r \in (0,1)$. Then

$$\begin{aligned}
\|r\,\mathbf{y} + (1-r)\,\mathbf{z} - \mathbf{x}\| \ &=\ \|r\,(\mathbf{y}-\mathbf{x}) + (1-r)\,(\mathbf{z}-\mathbf{x})\| \\
&\leq\ r\,\|\mathbf{y}-\mathbf{x}\| + (1-r)\,\|\mathbf{z}-\mathbf{x})\|, \qquad (\text{A5})
\end{aligned}$$

by the triangle inequality. Thus,

$$\begin{aligned}
\phi_\mathbf{x}\,(r\,\mathbf{y} + (1-r)\,\mathbf{z}) \ &=\ \phi\,(\|r\,\mathbf{y} + (1-r)\,\mathbf{z} - \mathbf{x}\|) \ \underset{(*)}{\leq}\ \phi\,(r\,\|\mathbf{y}-\mathbf{x}\| + (1-r)\,\|\mathbf{z}-\mathbf{x})\|) \\
&\underset{(\dagger)}{\leq}\ r\,\phi\,(\|\mathbf{y}-\mathbf{x}\|) + (1-r)\,\phi\,(\|\mathbf{z}-\mathbf{x})\|) \\
&=\ r\,\phi_\mathbf{x}(\mathbf{y}) + (1-r)\,\phi(\mathbf{z}), \qquad \text{as desired.} \qquad (\text{A6})
\end{aligned}$$

Here, $(*)$ is by inequality (A5), because ϕ is increasing, while $(\dagger)$ is because ϕ is convex, and becomes a strict inequality in the case when ϕ is strictly convex.

Now, for any $\mathbf{y} \in \mathbb{R}^N$, the defining equation (5) says $\Phi_\rho(\mathbf{y}) = \int_{\mathbb{R}^N} \phi_{\mathbf{x}}(\mathbf{y})\ \mathrm{d}\rho[\mathbf{x}]$. Thus, for any $\mathbf{y}, \mathbf{z} \in \mathbb{R}^N$, and any $r \in (0,1)$, we have

$$\begin{aligned}
\Phi_\rho\left(r\,\mathbf{y} + (1-r)\,\mathbf{z}\right) &= \int_{\mathbb{R}^N} \phi_{\mathbf{x}}\left(r\,\mathbf{y} + (1-r)\,\mathbf{z}\right)\ \mathrm{d}\rho[\mathbf{x}] \\
&\underset{(*)}{\leq} \int_{\mathbb{R}^N} r\,\phi_{\mathbf{x}}(\mathbf{y}) + (1-r)\,\phi_{\mathbf{x}}(\mathbf{z})\ \mathrm{d}\rho[\mathbf{x}] \\
&= r\int_{\mathbb{R}^N} \phi_{\mathbf{x}}(\mathbf{y})\ \mathrm{d}\rho[\mathbf{x}] \ + \ (1-r)\int_{\mathbb{R}^N} \phi_{\mathbf{x}}(\mathbf{z})\ \mathrm{d}\rho[\mathbf{x}] \\
&= r\,\Phi_\rho(\mathbf{y}) + (1-r)\,\Phi_\rho(\mathbf{z}),
\end{aligned}$$

as desired. Here, $(*)$ is by inequality (A6), and is a strict inequality in the case when ϕ is strictly convex. □

PROOF OF THEOREM 8. Recall that $\mathcal{A} \subset \mathbb{R}^N$. Let $\mathbf{a}, \mathbf{b} \in \mathcal{A}$. Let $\mathbf{v} := \mathbf{b} - \mathbf{a}$, and define:

$$\begin{aligned}
\mathcal{C}_{\mathbf{a}} &:= \left\{\mathbf{r} \in \mathbb{R}^N\ ;\ \|\mathbf{r}-\mathbf{a}\| < \|\mathbf{r}-\mathbf{b}\|\right\}, \\
\mathcal{H}_{\mathbf{a},\mathbf{b}} &:= \left\{\mathbf{r} \in \mathbb{R}^N\ ;\ \|\mathbf{r}-\mathbf{a}\| = \|\mathbf{r}-\mathbf{b}\|\right\}, \\
\text{and}\ \ \mathcal{C}_{\mathbf{b}} &:= \left\{\mathbf{r} \in \mathbb{R}^N\ ;\ \|\mathbf{r}-\mathbf{a}\| > \|\mathbf{r}-\mathbf{b}\|\right\}.
\end{aligned}$$

Then $\mathcal{C}_{\mathbf{a}}$ and $\mathcal{C}_{\mathbf{b}}$ are two open halfspaces in $\mathbb{R}^N$, separated by $\mathcal{H}_{\mathbf{a},\mathbf{b}}$, which is the hyperplane orthogonal to $\mathbf{v}$, and passing through the point $(\mathbf{a}+\mathbf{b})/2$.

CLAIM 1: *If $\mathbf{m}_\rho^\phi \in \mathcal{C}_{\mathbf{a}}$, then* $\lim_{I\to\infty}$ Prob $\Big($*A majority of $\{u_i\}_{i\in\mathcal{I}}$ prefer* $\mathbf{a}$ *over* $\mathbf{b}\Big) = 1$.

PROOF. Let $\mathcal{H}_{\mathbf{v}}^I \subset \mathbb{R}^N$ be any $\mathbf{v}$-median hyperplane of the collection $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ —that is, $\mathcal{H}_{\mathbf{v}}^I$ is a hyperplane in $\mathbb{R}^N$ orthogonal to $\mathbf{v}$, such that at least half the points in $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ lie either in $\mathcal{H}_{\mathbf{v}}^I$ or on one side of $\mathcal{H}_{\mathbf{v}}^I$, and at least half the points in $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ lie either in $\mathcal{H}_{\mathbf{v}}^I$ or on the other side of $\mathcal{H}_{\mathbf{v}}^I$. (Such a hyperplane may not be unique; if it is not unique, then just pick one arbitrarily.)

For any $i \in \mathcal{I}$, we have $u_i(\mathbf{a}) > u_i(\mathbf{b})$ if and only if $\mathbf{x}_i \in \mathcal{C}_{\mathbf{a}}$. It follows that

$$\Big(\text{A majority of } \{u_i\}_{i\in\mathcal{I}} \text{ prefer } \mathbf{a} \text{ over } \mathbf{b}\Big) \quad\Longleftrightarrow\quad \Big(\mathcal{H}_{\mathbf{v}}^I \subset \mathcal{C}_{\mathbf{a}}\Big). \tag{A7}$$

Let $\mathcal{H}_{\mathbf{v}}^\rho$ be the (unique) $\mathbf{v}$-median hyperplane of ρ; then condition (B4) says $\mathbf{m}_\rho^\phi \in \mathcal{H}_{\mathbf{v}}^\rho$. Thus, $\mathcal{H}_{\mathbf{v}}^\rho \subset \mathcal{C}_{\mathbf{a}}$ (because $\mathbf{m}_\rho^\phi \in \mathcal{C}_{\mathbf{a}}$ and $\mathcal{H}_{\mathbf{v}}^\rho$ is parallel to $\mathcal{H}_{\mathbf{a},\mathbf{b}}$). But as $I\to\infty$, the sample median hyperplane $\mathcal{H}_{\mathbf{v}}^I$ converges to $\mathcal{H}_{\mathbf{v}}^\rho$ in probability (by the Weak Law of Large Numbers). Thus, since $\mathcal{C}_{\mathbf{a}}$ is an open set containing $\mathcal{H}_{\mathbf{v}}^\rho$, we have

$$\lim_{I\to\infty} \text{Prob}\left[\mathcal{H}_{\mathbf{v}}^I \subset \mathcal{C}_{\mathbf{a}}\right] \quad=\quad 1. \tag{A8}$$

Combining statement (A7) with limit (A8) yields the claim. ◇ `Claim 1`

CLAIM 2: *There is a strictly increasing function* $\gamma : [0,\infty)\longrightarrow\mathbb{R}$ *such that* $\Phi_\rho(\mathbf{x}) = \gamma\left(\|\mathbf{x}-\mathbf{m}_\rho^\phi\|\right)$ *for all* $\mathbf{x}\in\mathbb{R}^N$.

PROOF. Condition (B3) implies that there is some function $\gamma : [0,\infty)\longrightarrow\mathbb{R}$ such that $\Phi_\rho(\mathbf{x}) = \gamma\left(\|\mathbf{x}-\mathbf{m}_\rho^\phi\|\right)$ for all $\mathbf{x}\in\mathbb{R}^N$. Lemma A1 says that Φ_ρ is convex; this implies that γ must be nondecreasing. Furthermore, the only place γ could fail to be strictly increasing (i.e. be constant) is in a neighbourhood of 0. But if γ was constant near 0, then Φ_ρ would be constant in a neighbourhood of $\mathbf{m}_\rho^\phi$, contradicting (B2). Thus, we conclude that γ is *strictly* increasing. ◇ Claim 2

Let $U_{\mathcal{I}} := \frac{1}{I}\sum_{i\in\mathcal{I}} u_i$, as in equation (1).

CLAIM 3: *If* $\mathbf{m}_\rho^\phi\in\mathcal{C}_{\mathbf{a}}$, *then* $\lim_{I\to\infty} \text{Prob}\left[U_{\mathcal{I}}(\mathbf{a}) > U_{\mathcal{I}}(\mathbf{b})\right] \ = \ 1.$

PROOF. Let γ be as in Claim 2. If $\mathbf{m}_\rho^\phi\in\mathcal{C}_{\mathbf{a}}$, then $\|\mathbf{a}-\mathbf{m}_\rho^\phi\| < \|\mathbf{b}-\mathbf{m}_\rho^\phi\|$; thus, $\Phi_\rho(\mathbf{a}) < \Phi_\rho(\mathbf{b})$ (because γ is strictly increasing). Fix $C\in\mathbb{R}$ with $\Phi_\rho(\mathbf{a}) < C < \Phi_\rho(\mathbf{b})$

Let $\mathbf{x}$ be a ρ-random variable. From equation (5) it is clear that $\Phi_\rho(\mathbf{a})$ is the expected value of $\phi(\|\mathbf{x}-\mathbf{a}\|)$. Meanwhile, $-U_{\mathcal{I}}(\mathbf{a}) = \frac{1}{I}\sum_{i\in\mathcal{I}}\phi(\|\mathbf{x}_i-\mathbf{a}\|)$ is an empirical estimate of this expected value, based on the sample set $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$. Thus, since $\Phi_\rho(\mathbf{a}) < C$, the Weak Law of Large Numbers says $\lim_{I\to\infty} \text{Prob}[-U_{\mathcal{I}}(\mathbf{a}) < C] = 1$. By a similar argument, $\lim_{I\to\infty} \text{Prob}[-U_{\mathcal{I}}(\mathbf{b}) > C] = 1$. Thus, $\lim_{I\to\infty} \text{Prob}\left[U_{\mathcal{I}}(\mathbf{a}) > -C > U_{\mathcal{I}}(\mathbf{b})\right] = 1$. ◇ Claim 3

If $\mathbf{m}_\rho^\phi\in\mathcal{C}_{\mathbf{a}}$, then Claims 1 and 3 together imply that

$$\lim_{I\to\infty} \text{Prob}\Big(\text{The utility profile } \{u_i\}_{i\in\mathcal{I}} \text{ is } \{a,b\}\text{-reasonable}\Big) \quad = \quad 1.$$

We can make a similar argument in the case when $\mathbf{m}_\rho^\phi\in\mathcal{C}_{\mathbf{b}}$. Finally, it is impossible that $\mathbf{m}_\rho^\phi\in\mathcal{H}_{\mathbf{a},\mathbf{b}}$, because $\|\mathbf{a}-\mathbf{m}_\rho^\phi\| \neq \|\mathbf{b}-\mathbf{m}_\rho^\phi\|$ by hypothesis.

This argument holds for any pair $a,b\in\mathcal{A}$. Since $\mathcal{A}$ is finite, we conclude that

$$\lim_{I\to\infty} \text{Prob}\Big(\text{The utility profile } \{u_i\}_{i\in\mathcal{I}} \text{ is reasonable}\Big) \ = \ 1. \qquad \square$$

Proposition 6 follows from Theorem 8 and the next result.

LEMMA A2. *Suppose* $\phi(x) = x^2$ *for all* $x\geq 0$. *Then any multivariate normal probability measure on* $\mathbb{R}^N$ *is* ϕ*-balanced.*

PROOF. Let ρ be a multivariate normal probablity measure on $\mathbb{R}^N$. As observed in the text, ρ is ϕ-balanced if and only if the mean of ρ lies in every median hyperplane of ρ. Let $\mathbf{v}\in\mathbb{R}^N$, and let ρ' be the orthogonal projection of ρ onto the line $\mathcal{L}$ through

$\mathbf{v}$. Then ρ' is also normal, and the mean of ρ' is just the orthogonal projection of the mean of ρ onto $\mathcal{L}$. Meanwhile, the $\mathbf{v}$-median hyperplane $\mathcal{H}^{\rho}_{\mathbf{v}}$ is just the hyperplane in $\mathbb{R}^N$ orthogonal to $\mathcal{L}$, passing through the median point of ρ'. But in a one-dimensional normal distribution, the mean equals the median. So the mean and median of ρ' are equal. This means that the mean of ρ lies in $\mathcal{H}^{\rho}_{\mathbf{v}}$, as desired. □

Proposition 7 follows from Theorem 8 and the next result.

PROPOSITION A3. *Let ρ be any probability measure on $\mathbb{R}$ which is symmetrically distributed about some point $\mathbf{m}$ in the support of ρ. Or, let $N \geq 2$, and let ρ be any probability measure on $\mathbb{R}^N$ which is rotationally symmetric around some point $\mathbf{m}$ in $\mathbb{R}^N$. Then for every strictly convex increasing function $\phi : [0,\infty)\longrightarrow\mathbb{R}$ satisfying inequality (4), the measure ρ is ϕ-balanced, with $\mathbf{m}^{\phi}_{\rho} = \mathbf{m}$.*

PROOF. First, note that ϕ satisfies inequality (4) if and only if it satisfies condition (B1). Thus, the function Φ_ρ is well-defined in equation (5).

CLAIM 1: $\mathbf{m}$ *is the unique global minimum of* Φ_ρ.

PROOF. First suppose $N \geq 2$. Since ϕ is strictly convex, Lemma A1 says that Φ_ρ is strictly convex. Thus, the global minimum of Φ_ρ is unique. But if ρ is rotationally symmetric around $\mathbf{m}$, then so is the function Φ_ρ. Thus, so is the set of global minima of Φ_ρ. Thus the (unique) global minimum must be at $\mathbf{m}$.

The argument in the case $N = 1$ is similar, except now "rotationally symmetric around $\mathbf{m}$" is changed to "symmetric under reflection across the point $\mathbf{m}$". ◇ Claim 1

CLAIM 2: *For every $\mathbf{v} \in \mathbb{R}^N$, the measure ρ has a unique $\mathbf{v}$-median hyperplane $\mathcal{H}^{\rho}_{\mathbf{v}}$, and $\mathbf{m} \in \mathcal{H}^{\rho}_{\mathbf{v}}$.*

PROOF. We will handle the cases $N = 1$ and $N \geq 2$ separately.

In the case $N = 1$, a median "hyperplane" is just a median point of ρ (the vector $\mathbf{v}$ is irrelevant in this case). The theorem hypothesis states that ρ is symmetrically distributed about $\mathbf{m}$. Thus, $\mathbf{m}$ is a median point of ρ. But we also assumed that $\mathbf{m}$ is in the support of ρ; thus, $\mathbf{m}$ is the *only* median point of ρ.

Now suppose $N \geq 2$. If ρ is rotationally symmetric around $\mathbf{m}$, then so is support(ρ). Thus, support(ρ) can be written as a union of concentric spheres centred at $\mathbf{m}$. Now let $\mathbf{v} \in \mathbb{R}^N$ be any vector, and define

$$\begin{aligned} \mathcal{C}^{-}_{\mathbf{v}} &:= \big\{\mathbf{r} \in \mathbb{R}^N \;;\; \mathbf{v} \bullet \mathbf{r} < \mathbf{v} \bullet \mathbf{m}\big\}, \\ \mathcal{H}^{\rho}_{\mathbf{v}} &:= \big\{\mathbf{r} \in \mathbb{R}^N \;;\; \mathbf{v} \bullet \mathbf{r} = \mathbf{v} \bullet \mathbf{m}\big\}, \\ \text{and } \mathcal{C}^{+}_{\mathbf{v}} &:= \big\{\mathbf{r} \in \mathbb{R}^N \;;\; \mathbf{v} \bullet \mathbf{r} > \mathbf{v} \bullet \mathbf{m}\big\}. \end{aligned}$$

Thus, $\mathcal{H}^{\rho}_{\mathbf{v}}$ is the unique hyperplane in $\mathbb{R}^N$ orthogonal to $\mathbf{v}$ and containing $\mathbf{m}$. Note that the halfspace $\mathcal{C}^{-}_{\mathbf{v}}$ can be transformed into $\mathcal{C}^{+}_{\mathbf{v}}$ by rotating 180 degrees through any axis passing through $\mathbf{m}$. Since ρ is rotationally symmetric around $\mathbf{m}$, this implies that $\rho[\mathcal{C}_{\mathbf{v}}^{-}] = \rho[\mathcal{C}_{\mathbf{v}}^{+}]$; thus, $\mathcal{H}_{\mathbf{v}}^{\rho}$ is a $\mathbf{v}$-median hyperplane for ρ. However, we have already noted that support(ρ) is a union of concentric spheres centred at $\mathbf{m}$; thus, $\mathcal{H}_{\mathbf{v}}$ intersects support(ρ). Thus, $\mathcal{H}_{\mathbf{v}}^{\rho}$ is the *only* $\mathbf{v}$-median hyperplane for ρ. This argument works for any $\mathbf{v} \in \mathbb{R}^N$. ◇ `Claim 2`

By hypothesis, ρ satisfies condition (B1). Claim 1 implies that ρ satisfies conditions (B2) and (B3), while Claim 2 implies that it satisfies condition (B4). Thus, ρ is Φ-balanced. □

References

Azrieli, Y. and S. Kim, "Pareto efficiency and weighted majority rules.," *International Economic Review* (to appear) (2014).

Badger, W. W., "Political Individualism, Positional Preferences, and Optimal Decision-Rules," in Niemi and Weisberg (1972), 34–59.

Bag, P. K., H. Sabourian and E. Winter, "Multi-stage voting, sequential elimination and Condorcet consistency," *J. Econom. Theory* 144 (2009), 1278–1299.

Bordley, R. F., "Using factions to estimate preference intensity: Improving upon one person/one vote," *Public Choice* 45 (1985), 257–268.

———, "One person/one vote is not efficient given information on factions," *Theory and Decision* 21 (1986), 231–250.

Curtis, R. B., "Decision-Rules and Collective Values in Constitutional Choice," in Niemi and Weisberg (1972), 23–33.

Dubey, P. and L. S. Shapley, "Mathematical properties of the Banzhaf power index," *Math. Oper. Res.* 4 (1979), 99–131.

Enelow, J. M. and M. J. Hinich, eds., *Advances in the Spatial Theory of Voting* (Cambridge, UK: Cambridge University Press, 2008).

Fleurbaey, M., "One stake one vote," (preprint), 2009.

Gelman, A., J. N. Katz and J. Bafumi, "Standard Voting Power Indexes Do Not Work: An Empirical Analysis," *British J. Polit. Sci* 34 (2004), 657–74.

Hinich, M. J. and M. C. Munger, *Analytical Politics* (Cambridge, UK: Cambridge University Press, 1997).

Kim, S., "Ordinal versus cardinal voting rules: a mechanism design approach," *(preprint)* (2014).

Laruelle, A. and F. Valenciano, "Egalitarianism and utilitarianism in committees of representatives," *Soc. Choice Welf.* 35 (2010), 221–243.

LEHTINEN, A., "The Welfare Consequences of Strategic Voting in Two Commonly used Parliamentary Agendas," *Theory and Decision* 63 (2007), 1–40.

———, "Strategic Voting and the Degree of Path-Dependence," *Group Decision and Negotiation* (to appear) (2014).

MCKELVEY, R. D., P. C. ORDESHOOK AND P. UNGAR, "Conditions for voting equilibria in continuous voter distributions," *SIAM J. Appl. Math.* 39 (1980), 161–168.

MILLER, N. R., "Graph-Theoretical Approaches to the Theory of Voting," *American Journal of Political Science* 21 (Nov 1977), 769–803.

NIEMI, R. G. AND H. F. WEISBERG, eds., *Probability Models of Collective Decision Making* (Columbus, Ohio: Charles E. Merrill Publishing, 1972).

NITZAN, S., *Collective Preference and Choice* (Cambridge University Press, 2009).

PIVATO, M., "Voting rules as statistical estimators," *Social Choice and Welfare* 40 (2013), 581–630.

———, "Asymptotic utilitarianism in scoring rules," *(preprint)* (2014a).

———, "Statistical Utilitarianism," *(preprint)* (2014b).

RAE, D., "Decision rules and individual values in constitutional choice," *Amer. Polit. Sci. Rev.* 63 (1969), 40–56.

SCHMITZ, P. W. AND T. TRÖGER, "The (sub-)optimality of the majority rule," *Games Econom. Behav.* 74 (2012), 651–665.

SCHOFIELD, N. J., "Is Majority Rule Special?," in Niemi and Weisberg (1972), 60–82.

STRAFFIN, P. D., JR., "Majority rule and general decision rules," *Theory and Decision* 8 (1977), 351–360.

TAYLOR, M., "Proof of a theorem on majority rule," *Behavioral Science* 14 (1969).