Thema Working Paper n°2014-18
Université de Cergy Pontoise, France

"Statistical utilitarianism"

Marcus Pivato

UNIVERSITÉ
de Cergy-Pontoise

CNRS
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

September, 2014

Statistical utilitarianism

Marcus Pivato[*]

THEMA, Université de Cergy-Pontoise[†]

and Department of Mathematics, Trent University[‡]

September 16, 2014

Abstract

Given a sufficiently large population satisfying certain statistical regularities, we show that it is often possible to accurately estimate the utilitarian social welfare function and identify the welfare-maximizing social alternative, even if we only have very noisy data about individual utility functions and interpersonal utility comparisons, and even if the individuals can be strategically dishonest.

Keywords: utilitarian; interpersonal comparisons; Groves-Clarke pivotal mechanism.

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

Utilitarianism may be philosophically attractive, but as a practical method for making collective decisions, it faces at least four major problems.

(Pr1) Interpersonal comparisons of utility are problematic. Even if we accept that such interpersonal comparisons are meaningful in principle, it is not clear how precise interpersonal comparisons could be made in practice.

(Pr2) It is difficult for the social planner to obtain accurate information about the voters' utility functions. (It is not generally feasible to obtain a precise utility assessment from every voter for every possible social alternative.)

(Pr3) Due to epistemic failures, a voter may have incorrect beliefs about the long-term consequences of various policy alternatives. Furthermore, people fail to correctly predict their own future utility level, even in apparently straightforward decision problems (Loewenstein and Schkade, 1999). Indeed, there is ample empirical evidence that people's beliefs about their own past, present, and future happiness are surprisingly unreliable, and subject to systematic biases, errors, and illusions (Kahneman et al., 1999). For instance, Sen (1985, Ch.1, p.15) suggests that poor or oppressed people sometimes become "habituated" to their miserable circumstances, and hence overstate their level of well-being. In short: voters do not even correctly perceive their *own* utility functions.

(Pr4) Voters may strategically misrepresent their utility functions, in order to manipulate the outcome.

However, in this paper, we will show that these problems almost disappear in large populations of voters satisfying certain mild statistical assumptions. In Section 2, we show that averaging utility data (even noisy or miscalibrated data) from a large population of voters will yield a good approximation of utilitarianism with high probability, despite problems (Pr1)-(Pr3). However, this result assumes that the utility-measurement errors associated with different voters are independent random variables. In Section 3, we show that the same conclusion holds for correlated random errors. In Section 4, we address the issue of strategic voting; we introduce a variant of the Groves-Clarke pivotal voting mechanism and show that it simultaneously solves problems (Pr1)-(Pr4) in a sufficiently large population.

Related literature. Lerner (1944, pp.29-32) was perhaps the first to deploy statistical aggregation to obviate the technical difficulties of utilitarianism. Under the plausible assumption that all people have diminishing marginal utility for wealth, Lerner argued that, even in a state of *total ignorance* about the precise structure of people's cardinal utility functions, an egalitarian distribution of wealth would maximize the *expected* aggregate utility for society, because the expected utility gains of the poor under such a wealth redistribution would more than cancel the expected utility losses of the rich. Lerner's original argument was obscure and generated much confusion; it was later clarified by Breit and

Culbertson Jr. (1970).[1] For Lerner's argument to work, his vague hypothesis of "total ignorance" about individual utility functions must be formalized in terms of quite specific assumptions about the probability distribution of the utility functions. The results of this paper can be seen as an extension of this approach.

2   Basic model

First we will fix some notation that will be maintained throughout the paper. Let $\mathbb{R}$ denote the set of real numbers. Let $\mathcal{A}$ denote a finite set of social alternatives, and let $A := |\mathcal{A}|$. Let $\mathcal{I}$ be a set of voters, and let $I := |\mathcal{I}|$. (We will typically suppose that I is very large.) For every i in $\mathcal{I}$, let $u_i : \mathcal{A} \longrightarrow \mathbb{R}$ be a 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 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 $U_{\mathcal{I}} : \mathcal{A} \longrightarrow \mathbb{R}$ defined by

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

Let $\operatorname{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$ denote the set of alternatives in $\mathcal{A}$ which maximize $U_{\mathcal{I}}$ —we will refer to these as *utilitarian optima*.[2] A utilitarian social planner wants to find a utilitarian optimum, but she may not have enough information to do this, because of the aforementioned problems (Pr1)-(Pr3). We will formalize her informational problems with two assumptions:

(U1) The interpersonal calibration constants $\{c_i\}_{i \in \mathcal{I}}$ are unknown. The social planner regards $\{c_i\}_{i \in \mathcal{I}}$ as independent (but not necessarily identically distributed) real-valued random variables.[3] There is a constant $\sigma_c^2 > 0$ such that, for every voter i in $\mathcal{I}$, the random variable c_i has a variance less than σ_c^2, and an expected value of 1.

(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 the function $v_i := u_i + \epsilon_i$, where $\epsilon_i : \mathcal{A} \longrightarrow \mathbb{R}$ is a random "error" term. For each alternative a in $\mathcal{A}$, the random errors $\{\epsilon_i(a)\}_{i \in \mathcal{I}}$ are independent.[4] They are not necessarily identically distributed, but they all have an expected value of 0, and 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 $\{\epsilon_i\}_{i \in \mathcal{I}}$.

---

[1]See also Lerner (1970), Breit and Culbertson Jr. (1972), McCain (1972), and McManus et al. (1972).

[2]Note that technically, $U_{\mathcal{I}}$ and its maximizer(s) depend on the specific profile of utility functions $\{u_i\}_{i \in \mathcal{I}}$. For simplicity, we have chosen not to indicate this explicitly in our notation.

[3]Presumably, $c_i > 0$ for all $i \in \mathcal{I}$. But we do not need to assume this to prove our results.

[4]Note 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 $\mathcal{A}$.

Assumption (U1) encodes problem (Pr1), while (U2) encodes both (Pr2) and (Pr3). 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.[5] If they *are* randomly generated, then we do *not* need to assume that $\{u_i\}_{i \in \mathcal{I}}$ are identically distributed, or assume that the random variables $u_i(a)$ and $u_i(b)$ are independent for a given voter i in $\mathcal{I}$ and distinct alternatives a and b in $\mathcal{A}$. However, we will assume the utility profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies two boundedness conditions:

(U3) There is a constant $M > 0$ such that $\dfrac{1}{I} \sum_{i \in \mathcal{I}} u_i(a)^2 < M^2$ for every a in $\mathcal{A}$.

(U4) There is a constant $\Delta > 0$ such that $\max_{\mathcal{A}}(U_{\mathcal{I}}) - U_{\mathcal{I}}(a) > \Delta$ for every $a \notin \mathrm{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$.[6]

The constant Δ in condition (U4) is the minimum "social welfare cost" of failing to choose a utilitarian optimum. Condition (U3) is clearly satisfied if $|u_i(a)| < M$ for every alternative a in $\mathcal{A}$ and every voter i in $\mathcal{I}$. Alternately, suppose that, for each a in $\mathcal{A}$, the utility values $\{u_i(a)\}_{i \in \mathcal{I}}$ are independent, identically distributed random variables drawn from a distribution with finite variance. If I is large, then the Law of Large Numbers says that (U3) will hold with very high probability.

For every alternative a in $\mathcal{A}$, we define the "observed" social welfare function

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

Thus, $V_{\mathcal{I}}$ is based on observable data (the functions $\{v_i\}_{i \in \mathcal{I}}$), and does not require the true values of the utility functions $\{u_i\}_{i \in \mathcal{I}}$ or the interpersonal calibration constants $\{c_i\}_{i \in \mathcal{I}}$. Therefore, the social planner can compute $V_{\mathcal{I}}$, and identify the alternatives in $\mathcal{A}$ which maximize $V_{\mathcal{I}}$. Our first result says that, if the population is sufficiently large, then $V_{\mathcal{I}}$ is a good approximation of $U_{\mathcal{I}}$, so that any social alternative which maximizes $V_{\mathcal{I}}$ will also maximize the value of $U_{\mathcal{I}}$, with very high probability.

Theorem 1 *For every voter i in $\mathcal{I}$, let $u_i : \mathcal{A} \longrightarrow \mathbb{R}$ be a utility function. Suppose that the profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies conditions (U3) and (U4), 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 (U1) and (U2). Define $U_{\mathcal{I}}$ and $V_{\mathcal{I}}$ as in equations (1) and (2). Then* $\lim_{I \to \infty} \mathrm{Prob} \left[\mathrm{argmax}_{\mathcal{A}}(V_{\mathcal{I}}) \subseteq \mathrm{argmax}_{\mathcal{A}}(U_{\mathcal{I}}) \right] = 1.$[7]

We can refine this result in three ways. First, we can drop condition (U4). Second, and relatedly, instead of demanding that a maximizer of $V_{\mathcal{I}}$ *exactly* maximizes $U_{\mathcal{I}}$, we could

---

[5]In this case, the results in this paper should be interpreted as statements which hold for any specific realization of these random utility functions.

[6]Most of our results involve taking a limit as $I \to \infty$. The constants M and Δ appearing in (U3) and (U4) are then assumed to be independent of I.

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

entertain 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 (1), then let

$$U_\mathcal{I}^* \quad := \quad \max\{U_\mathcal{I}(a) \; ; \; a \in \mathcal{A}\} \tag{3}$$

be the optimum utilitarian social welfare. Let $\delta > 0$ represent a "social suboptimality tolerance", and let $p > 0$ represent the probability that this tolerance will be exceeded (the planner wants both of these to be small). For any values of δ and p, we define

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

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

Theorem 2 *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* (U3). *Define $U_\mathcal{I}$, $V_\mathcal{I}$, and $U_\mathcal{I}^*$ as in equations (1), (2), and (3). For any $\delta > 0$ and any $p \in (0, 1)$, if $I \geq \overline{I}(\delta, p)$, then for any a_V^* in $\mathrm{argmax}_\mathcal{A}(V_\mathcal{I})$, we have $\mathrm{Prob}\left[U_\mathcal{I}(a_V^*) < U_\mathcal{I}^* - \delta\right] \; < \; p$.*

For example, suppose $|\mathcal{A}| = 6$, and for every alternative a in $\mathcal{A}$, suppose the utilities $\{u_i(a); \quad i \in \mathcal{I}\}$ are independent, uniformly distributed random variables ranging over some interval of length at most 10, contained within the interval $[-9, 9]$ (with perhaps different subintervals of $[-9, 9]$ for different alternatives in $\mathcal{A}$). Let $M := 5$; then for a large population of voters, condition (U3) will be satisfied with very high probability.[8] Suppose $\sigma_c^2 = 1$ and $\sigma_\epsilon^2 = 5$, and let $\delta := 0.2$ (i.e. 1.1% of the total utility range) and $p := 0.01$. Then $\overline{I}(\delta, p) = 1\,800\,000$. Thus, for a polity of two million voters, Theorem 1 says that, with 99% probability, the $V_\mathcal{I}$-maximal alternative will yield a $U_\mathcal{I}$-value within 1.1% of the theoretical optimum $U_\mathcal{I}^*$.

A remark on the informational assumptions. There are at least four methods by which the planner could acquire the data $\{v_i\}_{i\in\mathcal{I}}$ posited in assumption (U2). The first and most obvious method is through verbal surveys, interviews, or polls (e.g. with questions of the form, "On a scale of 1 to 100, how happy would you be with each of the following outcomes?"). To fix a quantitative scale, we might formulate such questions in terms of "willingness to pay" (e.g. "How much money would you pay to obtain each of the following outcomes?") or some other "willingness to sacrifice" (e.g. "What percent of your income would you give up?", "How many hours would you work?", "How many days would you go without food?"). Since utility is presumably a nonlinear function of these

---

[8] *Proof:* The variance of any such uniform distribution is at most $8\frac{1}{3} < 9$, and the square of its mean is at most $4^2 = 16$. Thus, its second moment will be at most $9 + 16 = 25 = 5^2$.

variables, we could get more accurate estimates by eliciting preferences over *lotteries* over these variables (assuming von Neumann-Morgenstern preferences). Of course, all of these methods are vulnerable to strategic misrepresentation (which is only partially mitigated if the sacrifices in question are actual and not merely hypothetical); we will address this issue in Section 4.

In addition to deliberate misrepresentation, surveys and other verbal methods may suffer from involuntary misreporting, due poor communication, cognitive biases, and/or problem (Pr3). So a second approach is to estimate an individual's utility by examining physiological data such as her blood pressure, muscle tension, pupil dilation, EEG, EKG and/or MRI patterns, and/or blood concentrations of dopamine, endorphin, epinephrine, and cortisol. Such data could be used to estimate people's *actual* utility levels (e.g. to detect correlations between utility and various socioeconomic, health, and/or lifestyle factors), but it could also be used to estimate an individual's utility for various *hypothetical* scenarios (e.g. by measuring her physiological responses to visual portrayals or verbal vignettes of these scenarios). This approach has been extensively investigated in the hedonic psychology literature (Kahneman et al., 1999, Part V), but it is still speculative, and unlikely to be feasible for a large population.

A third approach is to predict a population's utility response to hypothetical social alternatives with a mathematical model, perhaps informed by data collected using the two previous approaches. For example, suppose we had empirical data about the utility levels of a large population of individuals (e.g. estimated using verbal surveys and/or physiological measurements), along with putative causal variables such as their income, consumption of various private and public goods, health status, occupation, lifestyle, leisure activities, social situation, and place of residence, as well as fixed demographic variables such age, gender, family status, and ethnic background. For each value of the fixed demographic variable, we can construct the utility function (in terms of the putative causal variables) which best fits the empirical data.[9] We could then use this ensemble of utility functions to predict the utility response this population to a hypothetical perturbation of the status quo (e.g. a policy which shifts the equilibrium of the economy by some specified vector.) Such predictive utility models are indispensable when the decision has important long-term consequences, such as in environmental policy.

Finally, we can obtain utility information from individuals by asking them to vote. Although seemingly crude, this method can sometimes be surprisingly effective at estimating the utilitarian social welfare function (Pivato, 2014a,b).

Clearly, all of these methods are prone to considerable error, and provide little insight into interpersonal comparisons. This is the reason for the random variables $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i\}_{i \in \mathcal{I}}$ in assumptions (U1) and (U2). As should be clear from this informal discussion, we consider $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i\}_{i \in \mathcal{I}}$ as random variables *from the perspective of the social planner*, not from the perspective of some third party. We suppose that the planner has already obtained as much information as is feasible about the utility functions of the voters (e.g. through surveys, mathematical models, etc.), and that $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i\}_{i \in \mathcal{I}}$ represent the

---

[9]For this curve-fitting problem to be well defined, we would need to make some assumptions about the functional form of the utility function —e.g. that it is Cobb-Douglas, CES, logarithmic, etc.

residual uncertainty in her utility estimates, given all this information. Thus, Theorems 1 and 2 and the later theorems in this paper describe the *planner's* probability estimates, based on her information.

It is for this reason that we can assume without loss of generality that the random variables $\{\epsilon_i(a); i \in \mathcal{I}$ and $a \in \mathcal{A}\}$ all have zero expected value; if some $\epsilon_i(a)$ had nonzero expected value $\bar{\epsilon}$ (from the *planner's* perspective), then she could correct her estimate of $v_i(a)$ by adding $\bar{\epsilon}$ to it, leaving a residual error with zero expected value. Likewise, we can assume that $\{c_i\}_{i \in \mathcal{I}}$ all have expected value 1; if there was some i in $\mathcal{I}$ such that $\mathbb{E}[c_i] = \bar{c}_i \neq 1$, then the planner could replace v_i with $\widetilde{v}_i := \bar{c}_i \, v_i$, which we could regard as an estimate of the underlying utility function $\widetilde{u}_i := \bar{c}_i \, u_i$, with random error $\widetilde{\epsilon}_i := \bar{c}_i \, \epsilon_i$ (so that $\widetilde{v}_i = \widetilde{u}_i + \widetilde{\epsilon}_i$). If we replace c_i with $\widetilde{c}_i := c_i/\bar{c}_i$, then we would have $\widetilde{c}_i \, \widetilde{u}_i = c_i \, u_i$ (so the sum (1) would be unchanged), but $\mathbb{E}[\widetilde{c}_i] = 1$, as desired.

Of course, the social planner may be systematically *wrong* in her estimates. For example, due to some hidden bias or error in her methodology, she might systematically underestimate the preference intensities (as reflected in $\{c_i\}_{i \in \mathcal{I}}$) of some demographic group (e.g. low-income women). In this case, her estimates of the utilitarian social welfare function would be systematically incorrect. While this is possible, it is outside the scope of our model. We have assumed that the planner has already incorporated all information which is available to her. So if there is some remaining bias in her estimates, then this bias must be detectable only through information which is somehow *not* available to the planner —even in principle —yet which *is* available to some neutral third party. To be credible, a critique based on such "hidden bias" would need to explain how such information could remain hidden from the planner, yet be available to another observer.

Faster convergence. Theorem 2 shows that the population size I that is required to ensure a welfare loss less than δ with probability $1 - p$ has growth rate $\mathcal{O}(A/p\delta^2)$ as $p, \delta \to 0$ and $A \to \infty$. This estimate arises from Chebyshev's inequality. If we strengthen hypotheses (U1) and (U2), then we can obtain a bound on I that grows much more slowly. For example, suppose there are constants $C > 0$ and $E > 0$ such that $|c_i| \leq C$ and $|\epsilon_i(a)| < E$ for all $i \in \mathcal{I}$ and $a \in \mathcal{A}$. Let M be as in hypothesis (U3), and let $B := M \, C + E$. Then using Hoeffding's inequality in place of Chebyshev's, we get a variant of Theorem 1 where

$$\overline{I}(\delta, p) \quad = \quad \frac{8 \, B^2}{\delta^2} \log\left(\frac{2A}{p}\right). \tag{5}$$

For example, let $A := 6$, $M = 5$ and $\delta = 0.2$ as in the earlier example, and suppose $C = 1$ and $E = 5$. Then $\overline{I}(\delta, p) = 20\,000 \log(2A/p)$. Thus, for $p = 0.01$, we have $\overline{I}(\delta, p) \approx 142\,000$. So one hundred fifty thousand voters are enough to ensure a 99% probability that the $V_{\mathcal{I}}$-maximizing alternative has a $U_{\mathcal{I}}$-value within 1.1% of the theoretical optimum. Note that (5) grows much more slowly than (4) as $A \to \infty$ and $p \to 0$. We would also get a formula very similar to (5) if $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i(a)\}_{a \in \mathcal{A}}$ were independent normal random variables.

Almost-sure convergence. Theorems 1 and 2 follow from the fact that, for any $\delta > 0$,

$$\lim_{I \to \infty} \mathrm{Prob}\left[\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_\infty > \delta\right] \quad = \quad 0, \quad \text{where} \quad \|U_{\mathcal{I}} - V_{\mathcal{I}}\|_\infty := \max_{a \in \mathcal{A}} |U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)|.$$

This is a statement about convergence *in probability*. With stronger hypotheses, it is possible to establish a similar result for convergence *almost surely*. To do this, let $v_i := u_i + \epsilon_i$ as in assumption (U2), but now instead suppose:

(U2*) For each alternative a in $\mathcal{A}$, the random errors $\{\epsilon_i(a)\}_{i \in \mathcal{I}}$ are independent and *identically distributed*, with expected value zero (but possibly infinite variance).

An application of Kolmogorov's Strong Law of Large Numbers yields the following result.

Theorem 3 *Suppose that $c_i = 1$ for all $i \in \mathcal{I}$, while the random errors $\{\epsilon_i\}_{i \in \mathcal{I}}$ satisfy condition (U2*), and define $U_\mathcal{I}$ and $V_\mathcal{I}$ as in equations (1) and (2). Then for each $a \in \mathcal{A}$,*

$$\lim_{I \to \infty} \ \|U_\mathcal{I} - V_\mathcal{I}\|_\infty \quad = \quad 0, \quad \text{with probability 1.} \tag{6}$$

Let $U_\mathcal{I}^*$ be as in equation (3), and define $U_\mathcal{I}^\dagger := \min\{U_\mathcal{I}(a); \ a \in \operatorname*{argmax}_{\mathcal{A}}(V_\mathcal{I})\}$. Then

$$\lim_{I \to \infty} \ \left(U_\mathcal{I}^* - U_\mathcal{I}^\dagger\right) \quad = \quad 0, \qquad \text{with probability 1.} \tag{7}$$

In equation (7), the expression $U_\mathcal{I}^* - U_\mathcal{I}^\dagger$ is the maximum "social welfare loss" that the social planner risks if she chooses a social alternative which maximizes the observed social welfare function $V_\mathcal{I}$. The limit (7) says that, as the population size becomes very large, this maximum social welfare loss is almost certain to vanish.

However, Theorem 3 delivers too much and not enough at the same time. On the one hand, almost-sure convergence is actually *stronger* than the convergence in probability that we need in order to be confident of obtaining close-to-optimal policies. On the other hand, the "identically distributed" part of hypothesis (U2*) may be somewhat implausible, under some circumstances. Also, Theorem 3 assumes away problem (Pr1).[10] A third shortcoming of Theorem 3 is that it says nothing about the *speed* of convergence. If the population is "big enough", then $V_\mathcal{I}$ and $U_\mathcal{I}$ are essentially the same. But how big is "big enough"? Theorem 2 gives a precise answer to this question.[11]

3   Correlated voters

One obvious weakness of the model in Section 2 is the assumption that $\{c_i\}_{i \in \mathcal{I}}$ and $\{\epsilon_i\}_{i \in \mathcal{I}}$ are *independent* random variables. This may not be entirely realistic, because people

---

[10]To extend Theorem 3 to cover (Pr1), we must replace (U1) with the assumption (U1*): "The random variables $\{c_i\}_{i \in \mathcal{I}}$ are independent and *identically distributed*, and furthermore, their distribution is *symmetric* about 1." We must also replace (U3) with the assumption (U3′) from Section 3. We can then extend Theorem 3 to obtain the limits (6) and (7) under hypothesis (U1*), (U2*), and (U3′).

[11]We do have an *asymptotic* estimate on the rate of convergence in Theorem 3. The Law of the Iterated Logarithm says that, with probability 1, there is some I_0 such that $\|U_\mathcal{I} - V_\mathcal{I}\|_\infty < \sqrt{2 \log[\log |\mathcal{I}|]/|\mathcal{I}|}$ for all $\mathcal{I}$ with $|\mathcal{I}| \geq I_0$. However, I_0 itself is a random variable (it depends on the particular realization of the random error functions $\{\epsilon_i\}_{i \in \mathcal{I}}$), so it cannot be known in advance. Hence, for practical purposes, this asymptotic convergence rate is not very useful.

belonging to the same social group might have similar characteristics, leading to correlation amongst their preference intensities and measurement errors. For example, suppose we estimated people's utility functions using physiological data. It is quite possible that members of different demographic groups (e.g. different age groups, different genders, etc.) might exhibit, on average, different physiological responses for the same utility function. Alternately, suppose we tried to elicit utility data using a verbal questionnaire. Since members of different subcultures have slightly different understandings of the language, they may respond to the same survey questions in slightly different ways. Of course, a well-designed utility-measurement technology would recognize and correct for such group-specific biases. But it is still possible that some residual bias will remain, and this would lead to a correlation of measurement errors within each social group.

To obtain results comparable to Theorems 1 and 2, we need some constraints on the covariance of the random errors. For any $I \in \mathbb{N}$ and any $I \times I$ matrix $\mathbf{S} = [s_{i,j}]_{i,j=1}^{I}$, let

$$\|\mathbf{S}\|_1 \quad := \quad \sum_{i,j=1}^{I} |s_{i,j}|. \tag{8}$$

For all $I \in \mathbb{N}$, let $\mathcal{S}_I$ be a set of real-valued $I \times I$ matrices. (The covariance matrices for our random variables will be drawn from these sets.) We will assume that

$$\lim_{I \to \infty} \overline{\sigma}_I \quad = \quad 0, \tag{9}$$

where, for all $I \in \mathbb{N}$, we define $\overline{\sigma}_I := \sup\{\|\mathbf{S}\|_1 / I^2; \ \mathbf{S} \in \mathcal{S}_I\}$.

As in Section 2, we will suppose that, for each i in $\mathcal{I}$, the planner can only observe a function $v_i := u_i + \epsilon_i$, where $\epsilon_i : \mathcal{A} \longrightarrow \mathbb{R}$ is a random "error" term. For every alternative a in $\mathcal{A}$, let $\boldsymbol{\epsilon}(a) := [\epsilon_i(a)]_{i\in\mathcal{I}}$, an I-dimensional vector of these random errors. (The entries are not necessarily identically distributed.) Likewise, if $\{c_i\}_{i\in\mathcal{I}}$ are the interpersonal comparison coefficients from Section 2, then the planner regards $\mathbf{c} := [c_i]_{i\in\mathcal{I}}$ as another I-dimensional random vector. (Again, the entries are not necessarily identically distributed.) We now replace assumptions (U1)-(U3) with the following:

(**U1′**) The covariance matrix of $\mathbf{c}$ is an element of $\mathcal{S}_I$.[12]

(**U2′**) For every alternative a in $\mathcal{A}$, the covariance matrix of $\boldsymbol{\epsilon}(a)$ is an element of $\mathcal{S}_I$.

(**U3′**) There is some constant $M > 0$ such that $|u_i(a)| \leq M$ for every i in $\mathcal{I}$ and a in $\mathcal{A}$.

We must also supplement these with:

(**U0′**) For every a in $\mathcal{A}$, the covariance matrix of $\boldsymbol{\epsilon}(a)$ and $\mathbf{c}$ is an element of $\mathcal{S}_I$.[13]

---

[12]Recall: If X and Y are two real-valued random variables with expected values $\overline{X}$ and $\overline{Y}$, then their *covariance* is defined $\mathrm{cov}(X,Y) := \mathbb{E}\left[(X - \overline{X})(Y - \overline{Y})\right]$. If $\mathbf{X} = (X_1, \ldots, X_I)$ is an I-dimensional random vector, then its *covariance matrix* is the symmetric, positive definite, $I \times I$ matrix $\mathbf{S}$ where $s_{i,j} = \mathrm{cov}(X_i, X_j)$.

[13]If $\mathbf{X} = (X_1, \ldots, X_I)$ and $\mathbf{Y} = (Y_1, \ldots, Y_I)$ are I-dimensional random vectors, then $\mathrm{cov}(\mathbf{X}, \mathbf{Y})$ is the the $I \times I$ matrix $\mathbf{S}$ where $s_{i,j} = \mathrm{cov}(X_i, Y_j)$.

For example, if $\mathbf{c}$ and $\boldsymbol{\epsilon}(a)$ are independent (as we assumed in Section 2), then assumption (U0′) is satisfied as long as $\mathcal{S}_I$ contains the zero matrix. Likewise, assumptions (U1) and (U2) from Section 2 are just special cases of (U0′)-(U2′) (let $\mathcal{S}_I$ be the set of all non-negative diagonal matrices with entries bounded by $\max\{\sigma_c^2, \sigma_\epsilon^2\}$). Note that, like (U2), assumption (U2′) does *not* require the random error vectors $\boldsymbol{\epsilon}(a)$ and $\boldsymbol{\epsilon}(b)$ to be independent or identically distributed for $a \neq b$. The dependency structure between $\boldsymbol{\epsilon}(a)$ and $\boldsymbol{\epsilon}(b)$ is arbitrary. The covariance matrix of $\boldsymbol{\epsilon}(a)$ describes the pattern of correlations we expect to see in the society $\mathcal{I}$ with respect to utility-measurement errors concerning alternative a. We do *not* assume that all societies exhibit the same pattern of correlations for all social decision problems. Two societies of the same size may exhibit different patterns of correlations on the same social decision (e.g. because they have a different demographic structure, or a different patchwork of overlapping subcultures). The set $\mathcal{S}_I$ represents the set of *possible* patterns of correlations we could expect to see for any possible social decision, in any possible society of size I. For any $\mathbf{S} \in \mathcal{S}_I$, the ratio $\|\mathbf{S}\|_1 / I^2$ measures, roughly, the *average amount of correlation* between citizens in a society described by matrix $\mathbf{S}$.[14] Thus, the limit (9) says that the average correlation between citizens becomes smaller as their societies become larger. For example, suppose there are constants $J \in \mathbb{N}$ and $\varsigma > 0$ such that, for every $I \in \mathbb{N}$ and every $\mathbf{S} \in \mathcal{S}_I$, each row of $\mathbf{S}$ has at most J nonzero entries, and each has magnitude at most ς (i.e. any person in any society can be correlated with at most J other people, and each correlation can have strength at most ς). Then the limit (9) is satisfied. Or more generally, suppose there is some constant $K > 0$ such that the absolute sum of each row in $\mathbf{S}$ is at most K, for every $I \in \mathbb{N}$ and every $\mathbf{S} \in \mathcal{S}_I$ (i.e. any person can be correlated with as many other people as she wants, but she has a fixed "covariance budget" which she must distribute amongst them). Then again the limit (9) is satisfied. The next result says that the conclusion of Theorem 1 still holds in this setting.

Theorem 4 *For every voter i in $\mathcal{I}$, let $u_i : \mathcal{A} \longrightarrow \mathbb{R}$ be a utility function. Suppose that the profile $\{u_i\}_{i \in \mathcal{I}}$ satisfies conditions (U3′) and (U4), and suppose $\{c_i\}_{i \in \mathcal{I}}$, $\{\boldsymbol{\epsilon}(a)\}_{a \in \mathcal{A}}$ and $\{v_i\}_{i \in \mathcal{I}}$ are randomly generated according to rules (U0′)-(U2′). Define $U_{\mathcal{I}}$ and $V_{\mathcal{I}}$ as in equations (1) and (2). Then* $\lim\limits_{I \to \infty} \text{Prob}\left[\operatorname*{argmax}\limits_{\mathcal{A}}(V_{\mathcal{I}}) \subseteq \operatorname*{argmax}\limits_{\mathcal{A}}(U_{\mathcal{I}})\right] = 1.$[15]

As in Section 2, we can refine this result by computing the size of the population necessary to ensure that a $V_{\mathcal{I}}$-maximizer *almost* maximizes $U_{\mathcal{I}}$, with a some specified probability. Let M be as in assumption (U3′). For any $\delta > 0$ and any probability $p \in (0, 1)$, the limit equation (9) implies that, if I is large enough, then we have

$$\overline{\sigma}_I \quad \leq \quad \frac{p\,\delta^2}{4A\,(M+1)^2}. \tag{10}$$

---

[14]This is not exactly true, because the diagonal entries of the matrices in (U1′) and (U2′) are variances. But if I is large, then the sum of diagonals is dominated by the sum of the off-diagonal covariance terms.

[15]Here, as usual, the constants M and Δ in conditions (U3′) and (U4) are to be held constant as $I \to \infty$.

Let $\widetilde{I}(\delta, p)$ be the minimum value of I satisfying the inequality (10).[16] The next result plays a role similar to Theorem 2.

Theorem 5 *Suppose $\{u_i\}_{i \in \mathcal{I}}$, $\{c_i\}_{i \in \mathcal{I}}$, $\{\boldsymbol{\epsilon}(a)\}_{a \in \mathcal{A}}$ and $\{v_i\}_{i \in \mathcal{I}}$ satisfy (U0′)-(U3′). Define $U_{\mathcal{I}}$, $V_{\mathcal{I}}$, and $U_{\mathcal{I}}^*$, as in equations (1), (2), and (3). For any $\delta > 0$ and any $p \in (0,1)$, if $I \geq \widetilde{I}(\delta, p)$, then for any a_V^* in $\mathrm{argmax}_{\mathcal{A}}(V_{\mathcal{I}})$, we have $\mathrm{Prob}\left[U_{\mathcal{I}}(a_V^*) < U_{\mathcal{I}}^* - \delta\right] < p$.*

As we have noted, assumptions (U1) and (U2) are special cases of (U0′)-(U2′). Thus, if we replace (U3) with (U3′) in Theorems 1 and 2, then they become corollaries of Theorems 4 and 5. One possible shortcoming of Theorems 4 and 5 is that the boundedness assumption (U3′) is more restrictive than (U3). But if we strengthen (U1′) to (U1), and strengthen (U0′) to the assumption that $\mathbf{c}$ and $\boldsymbol{\epsilon}(a)$ are independent (as in Section 2), then we can weaken (U3′) to (U3) and still have the same conclusions.[17]

One might also worry that assumptions (U0′)-(U2′), while certainly much weaker than independence, still restrict the amount of correlation between the variables. An alternative approach, which avoids this concern, is to identify the elements of $\mathcal{I}$ with points on a grid (representing, e.g. geographic locations) and suppose that the random variables $\{\epsilon_i(a)\}_{i \in \mathcal{I}}$ form a stationary random field on this grid, for each $a \in \mathcal{A}$.[18] We must also assume that $\{c_i\}_{i \in \mathcal{I}}$ and $\{u_i\}_{i \in \mathcal{I}}$ are *independent* stationary random fields on this grid.[19] Then variants of Theorems 1 and 3 can be obtained, using the Birkhoff Ergodic Theorem in place of the Law of Large Numbers. However, it is more difficult to obtain a convergence speed result like Theorem 2 in this "grid" model. Furthermore, a grid is perhaps not the best way to model interpersonal correlations.

4   Strategyproof implementation

The results of Sections 2 and 3 address problems (Pr1)-(Pr3), but not (Pr4) —that is, the problem of strategic misrepresentation of utility functions. We will now present a mechanism which also deals with this last problem: a variant of the Groves-Clarke Pivotal Mechanism (Clarke, 1971; Groves, 1973).[20] To do this, we must make more specific assumptions about the voters' utility functions. We will suppose that each voter's utility is

---

[16]To provide a precise formula for $\widetilde{I}(\delta, p)$, analogous to formula (4), we would need more specific assumptions about the families $\mathcal{S}_I$ of admissible covariance matrices posited in assumptions (U0′)-(U2′). In the "covariance budget" example, we have $\overline{\sigma}_I = K/I$; thus we obtain $\widetilde{I}(\delta, p) = 4\,K A\,(M+1)^2/\delta^2\,p$.

If $(\mathbf{c}, \boldsymbol{\epsilon})$ was a multivariate normal random variable, then (10) could be replaced with the inequality $\overline{\sigma}_I \leq \delta^2/8\,(M+1)^2\,\ln(2A/p)$. Then we would have $\widetilde{I}(\delta, p) = 8\,K\,(M+1)^2\,\ln(2A/p)/\delta^2$ in the covariance budget example. So for normal random variables, $\widetilde{I}(\delta, p)$ grows very slowly as $A \to \infty$ and $p \to 0$.

[17]Indeed, in this case, we can redefine $\|\mathbf{S}\|_1$ as $\sum_{i,j=1}^{I} s_{i,j}$ in equation (8), which makes the limit condition (9) much easier to achieve.

[18]To be precise, the "grid" here is the group $\mathbb{Z}^2$. In fact, we could use any amenable group.

[19]Alternately, if $c_i = 1$ for all $i \in \mathcal{I}$, then $\{u_i\}_{i \in \mathcal{I}}$ could be arbitrary.

[20]Pivato (2013) develops another variant of the pivotal mechanism, which is in some ways similar to the one presented here. But the purpose of that variant was not to approximately implement utilitarianism; it was to be politically "equitable", by giving poor voters roughly the same influence as rich voters.

a function of both her net wealth and the social alternative which is chosen. For any voter i in $\mathcal{I}$, let $\widehat{u}_i : \mathcal{A} \longrightarrow \mathbb{R}$ be her cardinal utility function over the social alternatives, and let $u_i^\$: \mathbb{R} \longrightarrow \mathbb{R}$ be her cardinal utility over net wealth levels. We will assume that her combined utility function $U_i : \mathcal{A} \times \mathbb{R} \longrightarrow \mathbb{R}$ is *additively separable* —that is, $U_i(a, w_i) = \widehat{u}_i(a) + u_i^\$(w_i)$, for any social alternative a in $\mathcal{A}$ and any net wealth level w_i in $\mathbb{R}$. We also assume that the social alternative itself has no effect on the net wealth of the voters, so that the two coordinates a and w_i can be treated as independent variables.[21] Finally, we suppose that these utility functions admit one-for-one cardinal interpersonal comparisons.[22] Thus, we want to find the social alternative in $\mathcal{A}$ which maximizes the utilitarian social welfare function $\widehat{U}_\mathcal{I}$ defined by

$$\widehat{U}_\mathcal{I}(a) \quad := \quad \sum_{i \in \mathcal{I}} \widehat{u}_i(a) \quad \text{for all } a \text{ in } \mathcal{A}. \tag{11}$$

The problem is to elicit the true values of $\widehat{u}_i(a)$ from strategically dishonest voters. The Groves-Clarke Pivotal Mechanism solves this problem by asking each voter to declare the monetary price r she would be willing to pay (in the form of a so-called "Clarke tax") to obtain a given policy alternative a. The mechanism is designed such that it is a dominant strategy for each voter to reveal her true willingness-to-pay; thus, we can use $u_i^\$(r)$ to estimate the true value of $\widehat{u}_i(a)$.

However, the original Groves-Clarke mechanism has three major problems. First, it only works if voters have *quasilinear* utility functions —that is, if $u_i^\$(w_i)$ is a linear function of w_i. But it is much more likely that $u_i^\$$ is strictly concave (i.e. exhibits diminishing marginal utility). Second (and relatedly), by measuring utility via willingness-to-pay, the mechanism implements a weighted utilitarian SWF which gives the greatest weight to the voters having the lowest marginal utilities for wealth, who are generally the *richest* voters. So in effect, it implements a plutocracy.[23] Third, strategyproof mechanisms assume that all agents infallibly deploy their dominant strategies; it is not clear how the mechanism will perform in a world of fallible and boundedly rational agents.[24]

To extend the Groves-Clarke pivotal mechanism to the case when $u_i^\$$ is nonlinear, we will replace the deterministic Clarke tax of the original mechanism with a lottery. Formally, a *lottery* is an ordered pair $(p; r)$, where $0 \leq p \leq 1$ and $r \in \mathbb{R}$; this represents a random event which yields a payoff of r dollars with probability p, and a payoff of 0 dollars with probability $(1 - p)$. We will suppose that $u_i^\$$ represents i's von Neumann-Morgenstern (vNM) preferences on the set of such lotteries. Thus, if voter i's *ex ante* wealth level is w_i, then she will prefer lottery $(p; r)$ to lottery $(p'; r')$, if and only if $p\left(u_i^\$(w_i + r) - u_i^\$(w_i)\right) \geq p'\left(u_i^\$(w_i + r') - u_i^\$(w_i)\right)$. We will further suppose that each voter's joint utility over $\mathcal{A}$ and

---

[21]For the definition of "net wealth", and a discussion of these assumptions, see "Practicalities" below.

[22]That is, for any $a, b, c, d \in \mathcal{A}$ and $i, j \in \mathcal{I}$, if $\widehat{u}_i(b) - \widehat{u}_i(a) = \widehat{u}_j(d) - \widehat{u}_j(c)$, 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.

[23]A plutocracy may be Pareto efficient. But it is suboptimal in terms of the utilitarian SWF.

[24]A fourth problem is that the Clarke tax revenue cannot be redistributed to the voters, or the mechanism is no longer strategyproof. Instead, the Clarke tax revenue must be destroyed, which introduces an efficiency loss. But Green et al. (1976) and Green and Laffont (1979) showed that, under reasonable assumptions, the *per capita* inefficiency introduced by the Clarke tax goes to zero like $1/\sqrt{I}$ as $I \rightarrow \infty$. Thus, for the large-population asymptotic analysis in this paper, the Clarke tax inefficiency is negligible.

lotteries is *additively separable*. Thus, if alternative a is chosen and voter i must participate in the lottery $(p; r)$, then her *ex ante* utility will be $\widehat{u}_i(a) + p\,u_i^{\$}(w_i + r) + (1 - p)\,u_i^{\$}(w_i)$.

Suppose a_0 is the worst possible social alternative for voter i. By subtracting a constant from $\widehat{u}_i$ if necessary, we can assume without loss of generality that $\widehat{u}_i(a_0) = 0$. Let a be some other social alternative in $\mathcal{A}$. In our modified pivotal mechanism, instead of revealing $\widehat{u}_i(a)$ in terms of her "willingness to pay" for alternative a, voter i will reveal $\widehat{u}_i(a)$ by her "willingness to gamble". There will be a large price r that is fixed in advance. Voter i will be asked to name a probability p (between 0 and 1) such that she would be willing to participate in the lottery $(p, -r)$ in order to obtain alternative a rather than her worst-case alternative a_0. In our modified pivotal mechanism (see below), her dominant strategy will be to declare the unique probability ${}^{*}\!p_i^a$ in $[0, 1]$ such that

$$\widehat{u}_i(a) \quad = \quad {}^{*}\!p_i^a\, d_i, \tag{12}$$

where $d_i := u_i^{\$}(w_i) - u_i^{\$}(w_i - r)$ is the disutility that i would suffer from paying the price r. If we knew d_i, then we could compute $\widehat{u}_i(a)$ from ${}^{*}\!p_i^a$. By applying this approach for all voters and all alternatives, the social planner could accurately estimate the utilitarian social welfare function (11). However, there are still three caveats with this approach.

(i) The fixed price r must be large enough that $d_i \geq \widehat{u}_i(a)$, otherwise it is impossible to find a probability ${}^{*}\!p_i^a \in [0, 1]$ satisfying equation (12).

(ii) The social planner does not know the wealth-utility function $u_i^{\$}$, and there is no reliable way to elicit this information from voter i.

(iii) Due to cognitive errors, bounded rationality, and/or a poor understanding of her own utility function, voter i may make mistakes when computing or deploying her dominant strategy, and inadvertently fail to satisfy equation (12).

Each of these caveats necessitates a refinement of our approach. First consider Caveat (i). Note that equation (12) must be satisfied for every voter and every social alternative. In other words, r must be large enough that

$$u_i^{\$}(w_i) - u_i^{\$}(w_i - r) \quad \geq \quad \widehat{u}_i(a), \quad \text{for all } a \in \mathcal{A} \text{ and all } i \in \mathcal{I}. \tag{13}$$

However, in a society with large inequalities of wealth, a value of r large enough to satisfy (13) for very wealthy voters (who generally have very small marginal utilities for wealth) would be a terrifying or even unimaginable amount of money for very poor voters. This could introduce errors in the probability values p_i^a we obtain from the very poor voters. To deal with this, we will divide the population $\mathcal{I}$ into N subpopulations based on net wealth: $\mathcal{I} = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \cdots \sqcup \mathcal{I}_N$ (where $\mathcal{I}_1$ is the poorest group, and $\mathcal{I}_N$ is the richest in net wealth). We then specify an increasing sequence $r_1 < r_2 < \cdots < r_N$ of prices. Individuals in group $\mathcal{I}_n$ must gamble over lotteries with price r_n. For all $i \in \mathcal{I}$, let $n(i) \in [1 \ldots N]$ denote the group to which i belongs. The prices $r_1, \ldots, r_n$ must be large enough that

$$d_i \quad := \quad u_i^{\$}(w_i) - u_i^{\$}(w_i - r_{n(i)}) \quad \geq \quad \widehat{u}_i(a), \quad \text{for all } a \in \mathcal{A} \text{ and all } i \in \mathcal{I}. \tag{14}$$

However, Caveat (ii) means that we don't know d_i, so we cannot compute $\widehat{u}_i(a)$, even if we have obtained $^*p_i^a$ satisfying (12). Instead, we will approximate $u_i^\$$ by some "representative" wealth-utility function $u^\$$, which we will apply to all voters.[25] That is, for any voter i in $\mathcal{I}$, we will estimate the utility $\widehat{u}_i(a)$ by the value $u_i(a)$ defined by

$$u_i(a) \quad := \quad D_i \, {}^*p_i^a, \quad \text{where} \quad D_i \quad := \quad u^\$(w_i) - u^\$(w_i - r_{n(i)}). \tag{15}$$

This assumes that voter i will declare the probability $^*p_i^a$ defined by equation (12). But Caveat (iii) means that we cannot assume this, even if $^*p_i^a$ is her dominant strategy. Instead, we will suppose that she declares some probability $p_i^a = {}^*p_i^a + \epsilon_i^a$, where ϵ_i^a is a random error term. If we define $v_i(a) := D_i \, p_i^a$ and $\epsilon_i(a) := D_i \epsilon_a^i$, then we have $v_i(a) = u_i(a) + \epsilon_i(a)$, where u_i is as in equation (15).

Our modified pivotal mechanism now works as follows:

1. For every i in $\mathcal{I}$ and every a in $\mathcal{A}$, voter i must declare a probability value $p_i^a \in [0,1]$ for alternative a. We require $\min_{a \in \mathcal{A}} p_i^a = 0$.

2. For every i in $\mathcal{I}$ and a in $\mathcal{A}$, define $v_i(a) := D_i \, p_i^a$, where D_i is from equation (15).

3. For every alternative a in $\mathcal{A}$, define $\widehat{V}_{\mathcal{I}}(a) := \sum_{i \in \mathcal{I}} v_i(a)$.

4. The social choice a^* is the alternative in $\mathcal{A}$ which maximizes $\widehat{V}_{\mathcal{I}}$.

5. Suppose there is some i in $\mathcal{I}$ and some b in $\mathcal{A}$ such that $\widehat{V}_{\mathcal{I}}(a^*) - \widehat{V}_{\mathcal{I}}(b) \leq v_i(a^*) - v_i(b)$ (so i is a *pivotal* voter). Then i must participate in the lottery $(p_i, -r_{n(i)})$, where

$$p_i \quad := \quad \frac{1}{D_i} \sum_{j \in \mathcal{I} \setminus \{i\}} \Big(v_j(b) - v_j(a^*) \Big). \tag{16}$$

(Note that, by construction, $0 \leq p_i \leq 1$.)

We will say that this mechanism is *dominant-strategy truth-revealing* if, for every voter i in $\mathcal{I}$, her dominant strategy $^*p_i^a$ satisfies equation (12) for every a in $\mathcal{A}$. If i deploys this dominant strategy, then the function v_i defined in Step 2 will be equal to the function u_i from equation (15). By modifying the proof for the original Groves-Clarke pivotal mechanism, it is simple to show that the above mechanism is always dominant-strategy truth-revealing (see Theorem 6(a) below). But to overcome Caveats (i)-(iii), we will need some additional assumptions about $\{u_i^\$\}_{i \in \mathcal{I}}$.

($1) The wealth-utility functions $\{u_i^\$\}_{i \in \mathcal{I}}$ are independent random functions (not necessarily identically distributed). They are independent of $\{\epsilon_i\}_{i \in \mathcal{I}}$.

($2) The distributions of $\{u_i^\$\}_{i \in \mathcal{I}}$ are such that condition (14) is satisfied almost surely.

---

[25]The representative function $u^\$$ could be constructed by estimating the wealth-utility functions of some statistically representative sample of the population (e.g. using standard gambles) and then averaging.

($3) $\mathbb{E}\left[u_i^\$(w)\right] = u^\$(w)$, for all $i \in \mathcal{I}$ and all $w \in \mathbb{R}$.[26]

We also assume there are constants $C, K > 0$ such that, for all $i \in \mathcal{I}$:

($4) The variance of $u_i^\$(w_i) - u_i^\$(w_i - r_{n(i)})$ is less than $K \cdot \left(u^\$(w_i) - u^\$(w_i - r_{n(i)})\right)^2$.

($5) $u_i^\$(w_i) - u_i^\$(w_i - r_{n(i)}) > C \cdot \left(u^\$(w_i) - u^\$(w_i - r_{n(i)})\right)$, almost surely.

As in Section 2, assumption (1) means $\{u_i^\$\}_{i \in \mathcal{I}}$ are random *from the perspective of the social planner*; this encodes Caveat (ii). Assumption (2) addresses Caveat (i): it ensures that the prices $r_1, \ldots, r_N$ are large enough that all voters can adequately express the intensity of their preferences, while still satisfying the truth-revealing condition (12). Assumptions (3)-(5) say that the wealth-utility functions $\{u_i^\$\}_{i \in \mathcal{I}}$, while unknown, do not deviate too wildly from the "representative" wealth-utility function $u^\$$.

We now come to the main result of this section. It says that, if all voters in a large population deploy their dominant strategies in the mechanism (possibly plus some random error due to Caveat (iii)), then the outcome will maximize the utilitarian social welfare function $\widehat{U}_\mathcal{I}$ from equation (11), with very high probability. Thus, all of problems (Pr1)-(Pr4) can be solved simultaneously.

Theorem 6 *Let $\{\widehat{u}_i\}_{i \in \mathcal{I}}$ be arbitrary utility functions over $\mathcal{A}$, and let $\{u_i^\$\}_{i \in \mathcal{I}}$ be any set of wealth-utility functions satisfying condition (14) (e.g. via (2)). Then:*

(a) *The mechanism described by Steps 1 to 5 is dominant-strategy truth-revealing.*

Now suppose $\{u_i^\$\}_{i \in \mathcal{I}}$ satisfy conditions (1)-(5), while $\{\widehat{u}_i\}_{i \in \mathcal{I}}$ satisfy conditions (U3) and (U4) from Section 2.[27] For all $i \in \mathcal{I}$, let $v_i := u_i + \epsilon_i$, where u_i is from equation (15) and $\{\epsilon_i\}_{i \in \mathcal{I}}$ is a set of random error functions satisfying condition (U2).[28] Then:

(b) *If $\widehat{V}_\mathcal{I}$ is as in* Step 3 *of the mechanism, while $\widehat{U}_\mathcal{I}$ is as in equation (11), then*

$$\lim_{I \to \infty} \mathrm{Prob}\left[\operatorname*{argmax}_{\mathcal{A}}(\widehat{V}_\mathcal{I}) \subseteq \operatorname*{argmax}_{\mathcal{A}}(\widehat{U}_\mathcal{I})\right] = 1.$$

(c) *Let $\widehat{U}_\mathcal{I}^* := \max_\mathcal{A}(\widehat{U}_\mathcal{I})$. For any $\delta > 0$ and $p \in (0,1)$, let $\overline{I}(\delta, p)$ be as in equation (4). If $I \geq \overline{I}(\delta, p)$, then for any a_V^* in $\operatorname{argmax}_\mathcal{A}(\widehat{V}_\mathcal{I})$, $\mathrm{Prob}\left[\widehat{U}_\mathcal{I}(a_V^*) < \widehat{U}_\mathcal{I}^* - \delta\right] < p$.[29]*

If we replace assumption (U2) by assumption (U2′) from Section 3, then it is straightforward to obtain versions of Theorem 6(b,c) with correlated random errors, as in Theorems 4 and 5. We leave the details to the reader.

---

[26]In fact, we only need $\mathbb{E}[u_i^\$(w_i) - u_i^\$(w_i - r_{n(i)})] = u^\$(w_i) - u^\$(w_i - r_{n(i)})$, but it is simpler to state the more general condition.

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

[28]In other words, each voter deploys her dominant strategy from part (a), plus a random error.

[29]Part (c) actually does not require hypothesis (U4).

Practicalities. We have supposed that voter i's utility is a function both of the social alternative and her net wealth w_i. Here, by i's "net wealth", we mean the real value of $A_i - L_i + Y_i - X_i$, where $A_i = i$'s assets, $L_i = i$'s liabilities, $Y_i =$ expected net present value of i's future wages, and $X_i =$ expected net present value of i's future subsistence costs (e.g. minimal food, shelter and medical care for i and dependents). To construct the partition $\mathcal{I} = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \cdots \sqcup \mathcal{I}_N$, and to compute D_i in formula (15), the social planner must estimate w_i. In practice, the terms Y_i and X_i are impossible to compute. But for practical purposes, one could approximate $Y_i - X_i$ by $(y_i - x_i)/dr$, where y_i is i's *current* wages, x_i is *current* subsistence costs, and dr is the discount rate. The values of A_i, L_i, y_i and x_i can all be estimated fairly accurately using i's tax records; based on this, the social planner can estimate w_i.

We also assumed that voter i's joint utility function U_i is additively separable (i.e. $U_i(a, w_i) = \widehat{u}_i(a) + u_i^\$(w_i)$), and that the social alternatives have no effect on i's net wealth, so that a and w_i can be treated as independent variables. These assumptions are quite restrictive. If the social alternative a involve wealth transfers from rich to poor voters (either in cash, or in publicly provided goods and services), then U_i will generally not be separable. More subtly, many policy decisions shift the equilibrium price of goods in the economy. Thus, a could influence i's wealth-utility function, which again would violate separability. Similarly, a and w_i are not generally independent variables: if a involves taxes or cash transfers, then it will directly affect w_i. It could also affect w_i indirectly, by shifting the economic equilibrium, and thus, changing the market value of i's assets or labour, or the purchasing power of her money. (See Pivato (2013) for further discussion.)

These issues arise because we have used *money* as the yardstick for utility (via the stochastic Clarke tax). We could obviate them by denominating the Clarke tax in a non-monetary unit of value, such as hours of unpaid community service. But just as people have different marginal utilities for money, they have different marginal utilities for time, due to their differing opportunity costs. Thus, the mechanism would have to estimate these opportunity costs, by taking into account factors such as each voter's effective hourly wage and number of dependent family members.

5 Conclusion

According to conventional wisdom, utilitarianism may be appealing in theory, but is totally impossible to achieve in practice. The results of this paper suggest the opposite conclusion: utilitarianism is not merely possible, but actually fairly easy to achieve —at least as long as we are willing to tolerate a small amount of inefficiency, and as long as we have a sufficiently large population conforming to certain statistical regularities (e.g. hypotheses (U1), (U2), etc.). It is an empirical question whether real societies exhibit these regularities. If the results of these empirical investigations are affirmative, then "statistical" utilitarianism is indeed feasible. But is it desirable? This is a question of political philosophy.

Appendix: Proofs

Let $\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_{\infty} := \max\limits_{a \in \mathcal{A}} |U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)|$. Theorem 1 will follow from Theorem 2, which, in turn, will follow from the next result.

Proposition A1 *Let $U_{\mathcal{I}}$ and $V_{\mathcal{I}}$ be as in Theorem 2, and let $\delta > 0$ and $p \in (0, 1)$. If $I \geq \overline{I}(\delta, p)$, then* $\mathrm{Prob}\left[\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_{\infty} > \frac{\delta}{2}\right] < p$.

Proof. For all $a \in \mathcal{A}$, the quantity $U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)$ is a random variable. We will first compute its expected value and the variance of its distribution.

Claim 1: *For all $a \in \mathcal{A}$, $\mathbb{E}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] = 0$.*

Proof. For all $a \in \mathcal{A}$,

$$U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a) \underset{(\diamond)}{=\!=} \frac{1}{I}\sum_{i \in \mathcal{I}} c_i\, u_i(a) - \frac{1}{I}\sum_{i \in \mathcal{I}} v_i(a)$$

$$\underset{(\ddagger)}{=\!=\!=} \frac{1}{I}\sum_{i \in \mathcal{I}} c_i\, u_i(a) - \frac{1}{I}\sum_{i \in \mathcal{I}}\Big(u_i(a) + \epsilon_i(a)\Big)$$

$$= \frac{1}{I}\sum_{i \in \mathcal{I}}\Big((c_i - 1)\, u_i(a) - \epsilon_i(a)\Big), \qquad (A1)$$

where $(\diamond)$ is by defining equations (1) and (2), and $(\ddagger)$ is by assumption (U2). Thus,

$$\mathbb{E}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] = \frac{1}{I}\sum_{i \in \mathcal{I}} \mathbb{E}\Big[(c_i - 1)\, u_i(a) - \epsilon_i(a)\Big]$$

$$= \frac{1}{I}\sum_{i \in \mathcal{I}}\Big(\mathbb{E}[(c_i - 1)\, u_i(a)] - \mathbb{E}[\epsilon_i(a)]\Big)$$

$$\underset{(*)}{=\!=} \frac{1}{I}\sum_{i \in \mathcal{I}}\Big(u_i(a) \cdot \mathbb{E}[c_i - 1] + 0\Big) \underset{(\dagger)}{=\!=} \frac{1}{I}\sum_{i \in \mathcal{I}} u_i(a) \cdot 0 = 0.$$

as desired. Here, $(*)$ is because $u_i(a)$ is a constant and $\mathbb{E}[\epsilon_i(a)] = 0$ for all $i \in \mathcal{I}$, while $(\dagger)$ is because $\mathbb{E}[c_i] = 1$ for all $i \in \mathcal{I}$, by assumption (U1). $\diamond$ Claim 1

Claim 2: *For all $a \in \mathcal{A}$, $\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] \leq \dfrac{M^2 \cdot \sigma_c^2 + \sigma_\epsilon^2}{I}$.*

17

Proof. For all $a \in \mathcal{A}$,

$$
\begin{aligned}
\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] \;\;&\underset{\text{(a)}}{=\!=}\;\; \mathrm{var}\left[\frac{1}{I}\sum_{i \in \mathcal{I}}\Big((c_i - 1)\,u_i(a) - \epsilon_i(a)\Big)\right]\\[2mm]
&=\;\; \frac{1}{I^2}\mathrm{var}\left[\sum_{i \in \mathcal{I}}\Big((c_i - 1)\,u_i(a) - \epsilon_i(a)\Big)\right]\\[2mm]
&\underset{\text{(b)}}{=\!=}\;\; \frac{1}{I^2}\sum_{i \in \mathcal{I}}\Big(\mathrm{var}\left[(c_i - 1)\,u_i(a)\right] + \mathrm{var}\left[\epsilon_i(a)\right]\Big)\\[2mm]
&\underset{\text{(c)}}{=\!=}\;\; \frac{1}{I^2}\sum_{i \in \mathcal{I}}\Big(u_i(a)^2 \cdot \mathrm{var}\left[c_i\right] + \mathrm{var}\left[\epsilon_i(a)\right]\Big)\\[2mm]
&\underset{\text{(d)}}{\le}\;\; \frac{1}{I^2}\sum_{i \in \mathcal{I}}\Big(u_i(a)^2 \cdot \sigma_c^2 + \sigma_\epsilon^2\Big)\\[2mm]
&=\;\; \frac{\sigma_c^2}{I}\left(\frac{1}{I}\sum_{i \in \mathcal{I}}u_i(a)^2\right) + \frac{1}{I^2}\left(\sum_{i \in \mathcal{I}}\sigma_\epsilon^2\right) \quad\underset{\text{(e)}}{\le}\quad \frac{\sigma_c^2\,M^2}{I} + \frac{\sigma_\epsilon^2}{I},
\end{aligned}
$$

as desired. Here, (a) is by equation (A1), while (b) is because $\{\epsilon_i(a)\}_{i \in \mathcal{I}} \cup \{c_i\}_{i \in \mathcal{I}}$ are all jointly independent random variables, and (c) is because $\{u_i(a)\}_{i \in \mathcal{I}}$ are constants. Finally, (d) is by assumptions (U1) and (U2), while (e) is by (U3). $\quad\diamond$ Claim 2

Claim 3: *Let $q \in (0, 1)$, and suppose $I \ge \overline{I}(\delta, q)/A$. Then for all $a \in \mathcal{A}$, we have* $\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] < q$.

Proof. For any $a \in \mathcal{A}$, we have

$$
\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] \;\;\underset{(*)}{<}\;\; \frac{\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)]}{(\delta/2)^2} \;\;\underset{(\dagger)}{\le}\;\; 4\,\frac{M^2 \cdot \sigma_c^2 + \sigma_\epsilon^2}{I\,\delta^2} \;\;\underset{(\diamond)}{\le}\;\; q,
$$

as desired. Here, $(*)$ is by Claim 1 and Chebyshev's inequality, $(\dagger)$ is by Claim 2, and $(\diamond)$ is because $I \ge \overline{I}(\delta, q)/A = 4\,\dfrac{M^2 \cdot \sigma_c^2 + \sigma_\epsilon^2}{q\,\delta^2}$. $\quad\diamond$ Claim 3

Set $q = p/A$ in Claim 3. Then $\overline{I}(\delta, p) = \overline{I}(\delta, A\,q) = \overline{I}(\delta, q)/A$. Thus, if $I \ge \overline{I}(\delta, p)$, then for any particular $a \in \mathcal{A}$, Claim 3 says that $\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] \le q$. Since $A = |\mathcal{A}|$, this implies that

$$
\mathrm{Prob}\left[\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_\infty > \frac{\delta}{2}\right] = \mathrm{Prob}\left(|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2} \text{ for some } a \in \mathcal{A}\right) \le A \cdot q = p,
$$

as desired. $\qquad\square$

Remark. If the error vectors $\{\boldsymbol{\epsilon}(a)\}_{a \in \mathcal{A}}$ were all jointly independent, then the events $\{|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\}$ would be jointly independent (for all $a \in \mathcal{A}$). In this case, we would actually have $\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2} \text{ for some } a \in \mathcal{A}\right] \leq 1 - (1-q)^A$, which is a tighter estimate. From this, we could deduce that $\mathrm{Prob}\left[U_{\mathcal{I}}(a_V^*) < U_{\mathcal{I}}^* - \delta\right] < p$ whenever $I \geq \widehat{I}(\delta, p)$, where $\widehat{I}(\delta, p) := 4\,(M^2\,\sigma_c^2 + \sigma_\epsilon^2)/\delta^2\,(1 - (1 - \frac{p}{2})^{1/A})$. In general, $\widehat{I}(\delta, p) < \overline{I}(\delta, p)$. However, if p is close to zero, then $\widehat{I}(\delta, p)/\overline{I}(\delta, p) \approx 1$, so we do not obtain any significant improvement in practice from this much more restrictive independence hypothesis and the more complicated population bound it yields.

Proof of Theorem 2. Let $b^* \in \mathrm{argmax}_{\mathcal{A}}(U_{\mathcal{I}})$; thus, $U_{\mathcal{I}}(b^*) = U_{\mathcal{I}}^*$. Then

$$
\begin{aligned}
0 \ \leq \ U_{\mathcal{I}}^* - U_{\mathcal{I}}(a_V^*) \ &= \ U_{\mathcal{I}}(b^*) - U_{\mathcal{I}}(a_V^*) \\
&= \ U_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(b^*) + V_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(a_V^*) + V_{\mathcal{I}}(a_V^*) - U_{\mathcal{I}}(a_V^*) \\
&\leq \ U_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(b^*) \ + \ V_{\mathcal{I}}(a_V^*) - U_{\mathcal{I}}(a_V^*),
\end{aligned}
$$

where the last step is because $V_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(a_V^*) \leq 0$ because $a_V^* \in \mathrm{argmax}_{\mathcal{A}}(V_{\mathcal{I}})$. Thus, if $U_{\mathcal{I}}^* - U_{\mathcal{I}}(a_V^*) > \delta$, then either $U_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(b^*) > \frac{\delta}{2}$ or $V_{\mathcal{I}}(a_V^*) - U_{\mathcal{I}}(a_V^*) > \frac{\delta}{2}$. Thus,

$$
\begin{aligned}
\mathrm{Prob}\left[U_{\mathcal{I}}^* - U_{\mathcal{I}}(a_V^*) > \delta\right] \ &\leq \ \mathrm{Prob}\left(U_{\mathcal{I}}(b^*) - V_{\mathcal{I}}(b^*) > \tfrac{\delta}{2} \text{ or } V_{\mathcal{I}}(a_V^*) - U_{\mathcal{I}}(a_V^*) > \tfrac{\delta}{2}\right) \\
&\leq \ \mathrm{Prob}\left(|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \tfrac{\delta}{2} \text{ for some } a \in \mathcal{A}\right) \ < \ p,
\end{aligned}
$$

where the last inequality is by Proposition A1. $\qquad\square$

Proof of Theorem 1. Let Δ be as in assumption (U4). Let $\delta < \Delta$; then for any $a \in \mathcal{A}$, we have

$$
\left(U_{\mathcal{I}}(a) > U_{\mathcal{I}}^* - \delta\right) \iff \left(a \in \mathrm{argmax}_{\mathcal{A}}(U_{\mathcal{I}})\right).
$$

But Theorem 2 implies $\lim_{I \to \infty} \mathrm{Prob}\left(U_{\mathcal{I}}(a) > U_{\mathcal{I}}^* - \delta \text{ for all } a \in \mathrm{argmax}_{\mathcal{A}}(\mathcal{V}_{\mathcal{I}})\right) = 1$. $\qquad\square$

For any $\delta > 0$ and $p \in (0, 1)$, let $\widetilde{I}(\delta, p)$ be as defined below inequality (10). Theorems 4 and 5 are a consequence of the following result.

Proposition A2 *Let $U_{\mathcal{I}}$ and $V_{\mathcal{I}}$ be as in Theorem 5, and let $\delta > 0$ and $p \in (0, 1)$. If $I \geq \widetilde{I}(\delta, p)$, then $\mathrm{Prob}\left[\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_\infty > \frac{\delta}{2}\right] < p$.*

Proof. The proof strategy is very similar to Proposition A1. For all $a \in \mathcal{A}$, the quantity $U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)$ is a random variable. We will first compute its expected value and the variance of its distribution.

Claim 1: *For all $a \in \mathcal{A}$, $\mathbb{E}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] = 0$.*

Proof. Same as Claim 1 of Proposition A1. $\qquad\diamond$ Claim 1

19

Let $\overline{\sigma}_I$ be as defined below equation (9).

Claim 2:   *For all $a \in \mathcal{A}$, $\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] \leq (M+1)^2\,\overline{\sigma}_I$.*

Proof. Let $\mathbf{S}^c_{\mathcal{I}} := [s^c_{i,j}]_{i,j \in \mathcal{I}}$ be the $I \times I$ covariance matrix of the I-dimensional random vector $\mathbf{c}$, as described in assumption (U1'). Let $\mathbf{S}^a_{\mathcal{I}} := [s^a_{i,j}]_{i,j \in \mathcal{I}}$ be the $I \times I$ covariance matrix of the I-dimensional random error vector $\boldsymbol{\epsilon}(a)$, as described in assumption (U2'). Finally, let $\mathbf{S}^{c,a}_{\mathcal{I}} := [s^{c,a}_{i,j}]_{i,j \in \mathcal{I}}$ be the $I \times I$ covariance matrix of $\mathbf{c}$ and $\boldsymbol{\epsilon}(a)$, as described in assumption (U0'). Then

$$
\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)] \underset{\text{(a)}}{=\!=} \mathrm{var}\left[\frac{1}{I}\sum_{i \in \mathcal{I}}\left((c_i - 1)\,u_i(a) - \epsilon_i(a)\right)\right]
$$

$$
\underset{\text{(b)}}{=\!=} \frac{1}{I^2}\left(\sum_{i,j \in \mathcal{I}} u_i(a)^2\,s^c_{i,j} + 2\sum_{i,j \in \mathcal{I}} u_i(a)\,s^{c,a}_{i,j} + \sum_{i,j \in \mathcal{I}} s^a_{i,j}\right)
$$

$$
\underset{\text{(c)}}{\leq} \frac{1}{I^2}\left(\sum_{i,j \in \mathcal{I}} u_i(a)^2\,|s^c_{i,j}| + 2\sum_{i,j \in \mathcal{I}} |u_i(a)|\,|s^{c,a}_{i,j}| + \sum_{i,j \in \mathcal{I}} |s^a_{i,j}|\right)
$$

$$
\underset{\text{(d)}}{\leq} \frac{1}{I^2}\left(\sum_{i,j \in \mathcal{I}} M^2\,|s^c_{i,j}| + 2\sum_{i,j \in \mathcal{I}} M\,|s^{c,a}_{i,j}| + \sum_{i,j \in \mathcal{I}} |s^a_{i,j}|\right)
$$

$$
\underset{\text{(e)}}{=\!=} \frac{1}{I^2}\left(M^2\,\|\mathbf{S}^c\|_1 + 2M\,\|\mathbf{S}^{c,a}\|_1 + \|\mathbf{S}^a\|_1\right) \underset{\text{(f)}}{\leq} (M^2 + 2M + 1)\,\overline{\sigma}_I \;=\; (M+1)^2\,\overline{\sigma}_I,
$$

as desired. Here, (a) is by equation (A1) from the proof of Proposition A1. (b) is because $\{u_i(a)\}_{i \in \mathcal{I}}$ are all constants, and for any random variables X and Y and constant u, we have $\mathrm{var}(u\,X + Y) = u^2\,\mathrm{var}(X) + 2u\,\mathrm{cov}(X,Y) + \mathrm{var}(Y)$.[30] Next, (c) is by the triangle inequality, (d) is by hypothesis (U3'), (e) is by defining equation (8), and (f) is by the definition of $\overline{\sigma}_I$. $\diamond$ Claim 2

Claim 3:   *Let $q \in (0,1)$, and suppose $I \geq \widetilde{I}(\delta, A\,q)$. Then for all $a \in \mathcal{A}$, we have $\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] < q$.*[31]

Proof. For any $a \in \mathcal{A}$, we have

$$
\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] \underset{(*)}{<} \frac{\mathrm{var}[U_{\mathcal{I}}(a) - V_{\mathcal{I}}(a)]}{(\delta/2)^2} \underset{(\dagger)}{\leq} \frac{4\,(M+1)^2\,\overline{\sigma}_I}{\delta^2} \underset{(\diamond)}{\leq} q,
$$

as desired. Here, $(*)$ is by Claim 1 and Chebyshev's inequality, $(\dagger)$ is by Claim 2, and $(\diamond)$ is by inequality (10), because $I \geq \widetilde{I}(\delta, A\,q)$. $\diamond$ Claim 3

---

[30]If we strengthen (U1') to (U1), and strengthen (U0') to the assumption that $\mathbf{c}$ and $\boldsymbol{\epsilon}(a)$ are independent, while weakening (U3') to (U3), then this formula simplifies to $(M^2\,\sigma_c^2 + \sum_{i,j=1}^{I} s^a_{i,j})/I^2$. This leads to the variant of Theorems 4 and 5 described in footnote 17.

[31]If $(\mathbf{c}, \boldsymbol{\epsilon})$ was a multivariate normal random variable, then $V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)$ would also be normal; then Claim 3 would hold for the variant of inequality (10) that is mentioned in footnote 16.

Set $q = p/A$ in Claim 3. Then $\widetilde{I}(\delta, p) = \widetilde{I}(\delta, A\, q)$. Thus, if $I \geq \widetilde{I}(\delta, p)$, then for any $a \in \mathcal{A}$, Claim 3 says $\mathrm{Prob}\left[|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2}\right] \leq q$. Since $A = |\mathcal{A}|$, this implies that

$$\mathrm{Prob}\left[\|U_{\mathcal{I}} - V_{\mathcal{I}}\|_{\infty} > \frac{\delta}{2}\right] \;=\; \mathrm{Prob}\left(|V_{\mathcal{I}}(a) - U_{\mathcal{I}}(a)| > \frac{\delta}{2} \text{ for some } a \in \mathcal{A}\right) \;\leq\; A \cdot q \;=\; p,$$

as desired. $\qquad\square$

Proof of Theorems 4 and 5. Theorem 5 follows from Proposition A2 by an argument identical to the proof of Theorem 2. Then Theorem 4 follows from Theorem 5 by an argument identical to the proof of Theorem 1. $\qquad\square$

Proof of Theorem 6. (a) Let $\mathbf{p}_i := (p_i^a)_{a \in \mathcal{A}}$ be the vector of probabilities declared by voter i in Step 1 of the mechanism. We must show that $\mathbf{p}_i$ is a dominant strategy for i if and only if $p_i^a = {}^*p_i^a$ for all $a \in \mathcal{A}$, where ${}^*p_i^a$ is defined as in equation (12). The proof is similar to the analysis of the original Groves-Clarke pivotal mechanism, but with two important differences. First, in place of a deterministic Clarke tax, we now have a "stochastic" Clarke tax in Step 5 of the mechanism. Thus, voter i's payoff must be understood as her *expected* utility, prior to the resolution of the uncertainty in Step 5. Second, we must keep track of how the weight factors d_i and D_i affect i's dominant strategy, where d_i and D_i are defined by equations (14) and (15), respectively.

For any $a \in \mathcal{A}$, define $v_i(a) := D_i\, p_i^a$, as in Step 2 of the mechanism, and let $\widehat{V}_{-i}(a) := \sum_{j \in \mathcal{I} \setminus \{i\}} v_j(a)$. The vector $\mathbf{V}_{-i} := (\widehat{V}_{-i}(a))_{a \in \mathcal{A}}$ is an aggregate description of the behaviour of all other voters except i. The outcome of the mechanism (and thus, i's expected utility) are entirely determined by $\mathbf{p}_i$ and $\mathbf{V}_{-i}$. We will say that $\mathbf{p}_i$ is a *best response* for i to a given vector $\mathbf{V}$ if it maximizes i's expected utility when $\mathbf{V}_{-i} = \mathbf{V}$. For any $b \in \mathcal{A}$, let $\mathcal{V}_b := \{\mathbf{V} \in \mathbb{R}^{\mathcal{A}}; \; b = \mathrm{argmax}_{\mathcal{A}}(\mathbf{V})\}$.

Claim 1: *Fix $b \in \mathcal{A}$. Then $\mathbf{p}_i$ is a best response for i to all $\mathbf{V}_{-i} \in \mathcal{V}_b$ if and only if $p_i^{a_i} - p_i^b = (\widehat{u}_i(a_i) - \widehat{u}_i(b))/d_i$, where $a_i := \mathrm{argmax}_{\mathcal{A}}(v_i)$*

Proof. "$\Longrightarrow$" Suppose $\mathbf{V}_{-i} \in \mathcal{V}_b$; thus, $b = \mathrm{argmax}_{\mathcal{A}}(\mathbf{V}_{-i})$. Let's see how i's expected utility depends on the values of $p_i^{a_i}$ and p_i^b. There are two cases.

- If $p_i^{a_i} - p_i^b \geq [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$, then $v_i(a_i) - v_i(b) \geq \widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)$, so $\widehat{V}(a_i) > \widehat{V}(b)$, and hence $\widehat{V}(a_i) > \widehat{V}(c)$ for all $c \in \mathcal{A} \setminus \{a_i\}$. Thus, the winning alternative will be a_i in Step 4 of the mechanism, and i will pay the stochastic Clarke tax p_i given by formula (16) in Step 5. Thus, her expected utility will be $\widehat{u}_i(a_i) - d_i\,[\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$.

- On the other hand, if $p_i^{a_i} - p_i^b < [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$, then $v_i(a_i) - v_i(b) < \widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)$, so $\widehat{V}(b) > \widehat{V}(a_i)$ and hence $\widehat{V}(b) \geq \widehat{V}(c)$ for all $c \in \mathcal{A} \setminus \{b\}$. Thus, the winning alternative will be b in Step 4 of the mechanism, and i will pay nothing in Step 5. Thus, her expected utility will be $\widehat{u}_i(b)$.

21

It follows that a best response depends on whether or not $\widehat{u}_i(a_i) - d_i[\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i \geq \widehat{u}_i(b)$. There are two cases:

1. If $[\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i \geq [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$, then $\widehat{u}_i(a_i) - d_i [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i \geq \widehat{u}_i(b)$, so a best response $\mathbf{p}_i$ is such that $p_i^{a_i} - p_i^b \geq [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$.

2. If $[\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i < [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$, then $\widehat{u}_i(a_i) - d_i [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i < \widehat{u}_i(b)$, so a best response $\mathbf{p}_i$ is such that $p_i^{a_i} - p_i^b < [\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i$.

Suppose $\mathbf{p}_i$ is a best response for i to all $\mathbf{V}_{-i} \in \mathcal{V}_b$. Since Case 1 holds whenever $[\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i \leq [\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i$ we deduce that $\mathbf{p}_i$ must satisfy $p_i^{a_i} - p_i^b \geq [\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i$. Since Case 2 holds whenever $[\widehat{V}_{-i}(b) - \widehat{V}_{-i}(a_i)]/D_i > [\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i$, we deduce that $\mathbf{p}_i$ must satisfy $p_i^{a_i} - p_i^b \leq [\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i$. Combining these observations, we conclude that $\mathbf{p}_i$ must have $p_i^{a_i} - p_i^b = [\widehat{u}_i(a_i) - \widehat{u}_i(b)]/d_i$.

"$\Longleftarrow$" Suppose $\mathbf{p}_i$ is such that $p_i^{a_i} - p_i^b = (\widehat{u}_i(a_i) - \widehat{u}_i(b))/d_i$. By the analysis in the proof of "$\Longrightarrow$", we see that the strategy $\mathbf{p}_i$ always yields for i at least as good an expected utility as any other strategy, for any possible $\mathbf{V}_{-i} \in \mathcal{V}_b$. Thus, $\mathbf{p}_i$ is a best response for i to all $\mathbf{V}_{-i} \in \mathcal{V}_b$.　$\diamond$ Claim 1

The vector $\mathbf{p}_i$ is a dominant strategy for i if and only if it is a best response for i to *every* vector $\mathbf{V}_{-i} \in \mathbb{R}^{\mathcal{A}}$. By applying Claim 1 for all $b \in \mathcal{A}$, we deduce that $\mathbf{p}_i$ is a dominant strategy if and only if $p_i^{a_i} - p_i^b = (\widehat{u}_i(a_i) - \widehat{u}_i(b))/d_i$ for all $b \in \mathcal{A}$. By simple arithmetic, we deduce that $p_i^a - p_i^b = (\widehat{u}_i(a) - \widehat{u}_i(b))/d_i$ for all $a, b \in \mathcal{A}$.

But Step 1 of the mechanism requires that $\mathbf{p}_i$ must also satisfy $\min_{a \in \mathcal{A}} p_i^a = 0$. Meanwhile, we have assumed without loss of generality that $\min_{a \in \mathcal{A}} \widehat{u}_i(a) = 0$. From this, we deduce that $p_i^a = \widehat{u}_i(a)/d_i$ for all $a \in \mathcal{A}$. (Such a $\mathbf{p}_i$ is always possible, because condition (14) is satisfied by hypothesis.) In other words, $p_i^a = {}^*p_i^a$.

(b) Comparing equations (12) and (15) for all $i \in \mathcal{I}$, we see that $\widehat{u}_i = c_i\, u_i$, where $c_i := d_i/D_i$. If we substitute this into the utilitarian social welfare function in equation (11), we see that $\widehat{U}_{\mathcal{I}}/I = U_{\mathcal{I}}$, where $U_{\mathcal{I}}$ is the utilitarian social welfare function defined in equation (1). Thus, $\widehat{U}_{\mathcal{I}}/I$ satisfies assumption (U4) if and only if $U_{\mathcal{I}}$ satisfies (U4).

Claim 2:　*The set $\{c_i\}_{i \in \mathcal{I}}$ satisfies assumption (U1) from Section 2.*

Proof. For all $i \in \mathcal{I}$, equation (14) defines $d_i := u_i^{\$}(w_i) - u_i^{\$}(w_i - r_{n(i)})$. Hypothesis ($\$1$) implies that $\{d_i\}_{i \in \mathcal{I}}$ are independent random variables. Furthermore,

$$\mathbb{E}[d_i] \;=\; \mathbb{E}[u_i^{\$}(w_i) - u_i^{\$}(w_i - r_{n(i)})] \;\underset{(*)}{=\!=}\; u^{\$}(w_i) - u^{\$}(w_i - r_{n(i)}) \;\underset{(\dagger)}{=\!=}\; D_i,$$

for all $i \in \mathcal{I}$, where $(*)$ is by hypothesis ($\$3$), and $(\dagger)$ is by defining formula (15). Thus, if $c_i := d_i/D_i$, then $\{c_i\}_{i \in \mathcal{I}}$ are also independent random variables, such that $\mathbb{E}[c_i] = 1$ for all $i \in \mathcal{I}$. Hypothesis ($\$4$) implies that $\mathrm{var}[d_i] \leq K \cdot D_i^2$. Thus, $\mathrm{var}[c_i] = \mathrm{var}[d_i]/D_i^2 \leq K$. Thus, setting $\sigma_c^2 := K$, we see that $\{c_i\}_{i \in \mathcal{I}}$ satisfies (U1).　$\diamond$ Claim 2

Claim 3:　*The set $\{u_i\}_{i \in \mathcal{I}}$ defined by equation (15) satisfies condition (U3).*

Proof. By hypothesis, $\{\widehat{u}_i\}_{i\in\mathcal{I}}$ satisfy (U3); thus, there is some $\widehat{M} > 0$ such that, for all $a \in \mathcal{A}$, we have

$$\frac{1}{I}\sum_{i\in\mathcal{I}}\widehat{u}_i(a)^2 \quad < \quad \widehat{M}^2.$$

Hypothesis (\$5) implies $c_i > C$ for all $i \in \mathcal{I}$. Meanwhile, $u_i = \widehat{u}_i/c_i$ for all $i \in \mathcal{I}$. Thus,

$$\frac{1}{I}\sum_{i\in\mathcal{I}}u_i(a)^2 \quad = \quad \frac{1}{I}\sum_{i\in\mathcal{I}}\frac{\widehat{u}_i(a)^2}{c_i^2} \quad < \quad \frac{1}{I}\sum_{i\in\mathcal{I}}\frac{\widehat{u}_i(a)^2}{C^2} \quad < \quad \frac{\widehat{M}^2}{C^2},$$

for all $a \in \mathcal{A}$. Thus, the result follows by setting $M := \widehat{M}/C$. $\diamond$ Claim 3

Finally, the random error functions $\{\epsilon_i\}_{i\in\mathcal{I}}$ satisfy hypothesis (U2). Thus, if we define $V_I := \widehat{V}_I/I$, then Theorem 1 implies that

$$\lim_{I\to\infty}\text{Prob}\left[\operatorname*{argmax}_{\mathcal{A}}(V_{\mathcal{I}}) \subseteq \operatorname*{argmax}_{\mathcal{A}}(U_{\mathcal{I}})\right] \quad = \quad 1,$$

where $U_{\mathcal{I}}$ is from equation (1). But as we have observed, $U_{\mathcal{I}} = \widehat{U}_{\mathcal{I}}/I$.

Part (c) follows from Theorem 2 the same way that (b) followed from Theorem 1. $\square$

References

Breit, W., Culbertson Jr., W., June 1970. Distributional equality and aggregate utlity; comment. American Economic Review 60 (3), 435–41.

Breit, W., Culbertson Jr., W. P., June 1972. Distributional equality and aggregate utility: Reply. American Economic Review 62 (3), 501–502.

Clarke, E., Fall 1971. Multipart pricing of public goods. Public Choice 11, 17–33.

Green, J., Kohlberg, E., Laffont, J., 1976. Partial equilibrium approach to the free rider problem. Journal of Public Economics 6, 375–394.

Green, J., Laffont, J., 1979. Incentives in Public Decision Making. North-Holland, Amsterdam.

Groves, T., 1973. Incentives in teams. Econometrica 41, 617–631.

Kahneman, D., Diener, E., Schwarz, N. (Eds.), 1999. Well-being: The foundations of hedonic psychology. Russell Sage Foundation, New York, NY.

Lerner, A. P., 1944. The Economics of Control. New York.

Lerner, A. P., June 1970. Distributional equality and aggregate utility: Reply. American Economic Review 60 (3), 442–43.

Loewenstein, G., Schkade, D., 1999. Wouldn't it be nice? predicting future feelings. In: Kahneman et al. (1999), Ch. 5, pp. 85–105.

McCain, R., June 1972. Distributional equality and aggregate utility: Further comments. American Economic Review 62 (3), 497–500.

McManus, M., Walton, G. M., Coffman, R. B., June 1972. Distributional equality and aggregate utility: Further comment. American Economic Review 62 (3), 489–496.

Pivato, M., 2013. An equitable pivotal mechanism. (preprint).

Pivato, M., 2014a. Asymptotic utilitarianism in scoring rules. (preprint).

Pivato, M., 2014b. Condorcet meets Bentham. (preprint).

Sen, A., 1999. Commodities and capabilities. Oxford Univ. Press, New Delhi.