INTERNATIONAL ECONOMIC REVIEW
Vol. 55, No. 4, November 2014

PARETO EFFICIENCY AND WEIGHTED MAJORITY RULES*

BY YARON AZRIELI AND SEMIN KIM[1]

The Ohio State University, U.S.A.; National Taiwan University

We consider the design of decision rules in an environment with two alternatives, independent private values and no monetary transfers. The utilitarian rule subject to incentive compatibility constraints is a weighted majority rule, where agents' weights correspond to *expected* gains given that their favorite alternative is chosen. It is shown that a rule is interim incentive efficient if and only if it is a weighted majority rule, and we characterize those weighted majority rules that are ex ante incentive efficient. We also discuss efficiency in the class of anonymous mechanisms and the stability of weighted majority rules.

1. INTRODUCTION

Which mechanisms generate economically efficient outcomes when agents hold relevant private information? This is one of the most basic and important questions analyzed by the extensive mechanism design literature of the last several decades. In this article, we address this question in what is probably the simplest nontrivial environment: Society needs to choose between two alternatives, say *reform* or *status quo*. Each agent is privately informed about his utility for each of the alternatives. The uncertainty of agents about the types of their opponents is captured by a common prior distribution, which we assume to be independent across agents. Monetary transfers are prohibited.

The simplicity of the environment enables us to get a clear and precise answer to the above basic question by characterizing the class of efficient mechanisms subject to incentive compatibility. The characterization depends on the particular notion of efficiency considered, but in all cases we find that the family of *weighted majority rules* plays an important role. These voting mechanisms, which include the simple majority rule as a special case, are widely used in practice and have been extensively studied in the political and economic literature.[2] We think of our results as providing a normative rationale for the use of weighted majority rules, one that is based on the classic notions of efficiency and incentive compatibility.

Let us be more explicit regarding what we call a weighted majority rule. A Social Choice Function (SCF) f is a mapping from type profiles to lotteries over {reform, status-quo}. We say that f is a weighted majority rule if we can find positive weights $(w_1, \ldots, w_n)$ for the agents and a positive quota q such that under f the reform is implemented if the sum of weights of agents that prefer the reform exceeds q, and the status-quo prevails if this sum is less than q. Ties are allowed to be broken in an arbitrary way.

Our first result (Theorem 1) characterizes SCFs that maximize (ex ante, utilitarian) social welfare subject to incentive compatibility.[3] We show that an SCF is a solution to this optimization

*Manuscript received July 2012; revised July 2013.

[1] The authors thank Miguel Ballester, Tilman Börgers, PJ Healy, Dan Levin, Dipjyoti Majumdar, Jim Peck, and Arunava Sen for helpful comments and references. We are also grateful to three anonymous referees and to the editor Hülya Eraslan for their feedback and suggestions that improved the quality of this article. Please address correspondence to: Yaron Azrieli, Department of Economics, The Ohio State University, 465B Arps Hall, 1945 North High Street, Columbus, OH 43210. E-mail: *azrieli.2@osu.edu*

[2] See Section 2 for references.

[3] By the revelation principle, and assuming that agents play Bayes–Nash equilibrium, incentive compatibility constraints exactly characterize the class of feasible SCFs.

problem if and only if it is a weighted majority rule with specific weights. In particular, incentive compatibility constraints prevent effective use of any information about *realized* intensity of preferences; only the ordinal ranking of the two alternatives as reported by the agents matters for the outcome. However, the *distribution* of intensities is relevant for determining the optimal rule, since, roughly speaking, the weight assigned to agent i in the optimal rule reflects the *expected* utility gain of i if his favorite alternative is chosen.

We then move on to characterize the classes of interim and ex ante incentive efficient SCFs (Theorems 2 and 3). First, we show that an SCF is interim incentive efficient if and only if it is a weighted majority rule. In other words, the Pareto frontier (in the interim stage) of the set of incentive compatible (IC) decision rules coincides with the class of weighted majority rules. Second, we characterize the subset of weighted majority rules that are also ex ante incentive efficient by imposing restrictions on their weights and quota. Thus, in environments that are likely to satisfy our assumptions, the only kind of decision schemes that should be used are weighted majority rules. Indeed, any other IC rule is Pareto inferior to some (IC) weighted majority rule. These results are proved by showing that an SCF is (interim or ex ante) incentive efficient in a given environment if and only if it is a maximizer of social welfare in an auxiliary "equivalent" environment, and then applying Theorem 1.

Although the environment we analyze is simple, we emphasize that no symmetry between the agents or the alternatives is assumed. Even at the ex ante stage, different agents may have different utility distributions, and these distributions may be biased in favor of one of the alternatives. We believe that this is an important feature of our analysis, since many real-life environments are inherently asymmetric. Examples include representative democracies with heterogenous district sizes, publicly held firms with institutional and private shareholders, and faculty hiring decisions in which the job candidate has research interests closer to some faculty members than others. We discuss the implications of our results for some of these environments in Section 8.

Even if there is asymmetry between the agents, fairness considerations may force the designer to consider only *anonymous* mechanisms. Anonymity means that the chosen alternative depends only on the numbers of agents who prefer each of the alternatives and not on their names. In Section 6, we characterize the utilitarian solution and efficient rules under the additional constraint that only anonymous SCFs are feasible. As can be expected, *qualified majority rules*—which are weighted majority rules in which all the agents have the same weight—become central to the analysis. Specifically, we show that the utilitarian rule is a qualified majority rule with a certain quota (Theorem 4), that qualified majority rules are exactly those anonymous SCFs that are interim incentive efficient (Theorem 5), and that qualified majority rules with certain range of quotas are ex ante incentive efficient (Theorem 6).

Our last theoretical contribution in this article is a preliminary analysis of the *stability* of weighted majority rules (Section 7). Stability refers to a rule being sufficiently popular so that no other rule will get strong enough support to replace the existing rule. Several formulations of this idea appear in the literature, the differences between them being the timing at which the agents need to choose between the rules (interim or ex ante) and the majority that is needed in order to replace the incumbent rule by the challenger (unanimity or the majority implied by the incumbent rule itself). We show that the concept of ex ante self-stability (decision at the ex ante stage by the current rule) is very restrictive, namely, that in symmetric environments no qualified majority rule is ex ante self-stable. On the other hand, we show that durability (decision at the interim stage by unanimity) is very permissive in our environment, since every weighted majority rule is durable. We leave a more comprehensive analysis of the stability of weighted majority rules for future research.

The environment we study has very unique features (two alternatives, independent private values, no transfers), and as we demonstrate in Section 9 our results do not extend to more general environments. However, from a practical point of view, the two-alternatives case is perhaps the most interesting one to study, since binary decision problems are frequent. The no transfers assumption is also realistic, since in many cases they are infeasible or excluded for ethical reasons. Type independence is more restrictive in our context, but given the pervasiveness of this assumption in the literature we think that it is an interesting benchmark to study.

2. RELATED LITERATURE

The evaluation of (anonymous) majority rules based on the ex ante expected welfare they generate first appears in Rae (1969), and the analysis was later extended in Badger (1972) and Curtis (1972). A more recent treatment appears in Schmitz and Tröger (2011; see also the working paper Schmitz and Tröger, 2006), which analyzes efficiency of anonymous SCFs in a symmetric (between the agents and the alternatives) environment. Under incentive compatibility, the authors show that simple majority is the utilitarian rule (Proposition 2) and that it interim Pareto dominates every other IC, anonymous, and neutral rule (Proposition 3).[4] Our results complement these findings by extending the analysis to asymmetric environments and by considering both cases of anonymous and nonanonymous rules.

The theoretical study of weighted majority rules (with heterogenous weights) dates back at least to von-Neumann and Morgenstern's book (1944, Section 5). Much of the literature deals with the measurement of the power of players in the cooperative game generated by the rule (e.g., Shapley and Shubik, 1954). In Barberà and Jackson (2006), a model of indirect democracy is studied, where each country in a union has a single representative that votes on its behalf. The authors find the optimal weights (according to a utilitarian criterion) that should be assigned to the representatives. These weights are essentially the same as in our Theorem 1. We discuss in greater detail the connection between the papers in Section 8. A related analysis of the weights of representatives of heterogenous districts is done in Beisbart and Bovens (2007).

Another result closely related to our Theorem 1 is Proposition 2 in Fleurbaey (2008). The main difference is that we derive the optimality of the weighted majority rule among all IC SCFs, whereas Fleurbaey (2008) considers only ordinal SCFs. On the other hand, Fleurbaey (2008) allows for correlated types.

The voting example in Nehring (2004, Section 2) is also related to our analysis. In the same environment as in this article, the author observes that a specific weighted majority rule maximizes ex ante social welfare and that there are many weighted majority rules that are interim incentive efficient. The analysis in Nehring (2004) presumes that incentive compatibility is equivalent to strategy-proofness, but these two concepts differ even in this simple environment, and it is easy to construct examples of SCFs that are IC but not strategy-proof.[5]

When there are more than two alternatives the analysis of efficient SCFs becomes much more involved. In Apesteguia et al. (2011), the authors characterize the utilitarian rule among the SCFs that use only ordinal information, and show that it has the form of a scoring rule. They are not concerned with incentive compatibility issues, but Kim (2012) proves that the rule characterized in Apesteguia et al. (2011) is in fact IC in neutral environments. However, Kim (2012) also shows that there is an IC SCF that uses cardinal information and yields higher expected welfare than any ordinal SCF.

Our analysis of stability in Section 7 borrows concepts from earlier works on this topic. Namely, ex ante self-stability for qualified majority rules is introduced and analyzed in Barberà and Jackson (2004). The durability concept is due to Holmström and Myerson (1983).

There are several papers that consider the design of mechanisms in other environments when monetary payoffs are not allowed. In Börgers and Postl (2009), the authors consider a two-agent model with privately known intensities of preferences over three alternatives and derive results

[4] The analysis in Schmitz and Tröger (2011) also covers the case where incentive compatibility is replaced by the stronger strategy-proofness requirement. The results in this case are stronger and in particular allow for correlated types.

[5] With two alternatives, incentive compatibility is equivalent to ordinal incentive compatibility introduced in Majumdar and Sen (2004). With at least three alternatives, however, ordinal incentive compatibility is much stronger than incentive compatibility.

regarding the form of the second-best decision rule. Another example of Bayesian mechanism design without transfers is Miralles (2012), who considers the allocation of two objects among n agents. In Gershkov et al. (2013), optimal (utilitarian) dominant-strategy mechanisms in an environment with linear utilities but without transfers are characterized. Finally, our interest in the current question was initiated by Example 1 in Jackson and Sonnenschein (2007, pp. 243–44), which shows that incentive constraints can be overcome by linking many decision problems.

3. ENVIRONMENT

We consider a standard Bayesian environment à la Harsanyi. The set of agents is $N = \{1, 2, \ldots, n\}$ with $n \geq 1$. For each $i \in N$, T_i is a finite set of possible types of agent i, and t_i denotes a typical element of T_i. The type of agent i is a random variable $\hat{t}_i$ with values in T_i. The distribution of $\hat{t}_i$ is[6] $\mu_i \in \Delta(T_i)$, which we assume has full support. Let $T = T_1 \times \cdots \times T_n$ be the set of type profiles. We assume that types are independent across agents, so the distribution of $\hat{t} = (\hat{t}_1, \ldots, \hat{t}_n)$ is the product distribution $\mu = \mu_1 \times \cdots \times \mu_n \in \Delta(T)$. As usual, a subscript $-i$ means that the ith coordinate of a vector is excluded.

Let $A = \{\text{reform, status-quo}\} = \{r, s\}$ be the set of alternatives. The utility of each agent depends on the chosen alternative and on his own type only (private values). Specifically, the utility of agent i is given by the function $u_i : T_i \times A \to \mathbb{R}$. For ease of notation, we write $u_i^r(t_i) = u_i(t_i, r)$, $u_i^s(t_i) = u_i(t_i, s)$, and $u_i(t_i) = (u_i^r(t_i), u_i^s(t_i))$. For expositional purposes, we assume that no agent is ever indifferent between the two alternatives, that is $u_i^r(t_i) \neq u_i^s(t_i)$ for every $t_i \in T_i$ and every $i \in N$. In addition, we assume that for every i there are $t_i, t_i' \in T_i$ such that $u_i^r(t_i) > u_i^s(t_i)$ and $u_i^r(t_i') < u_i^s(t_i')$.

Since randomization over alternatives will be considered, we need to extend each $u_i(t_i, \cdot)$ to $\Delta(A)$. We identify $\Delta(A)$ with the interval $\{(p, 1-p) : 0 \leq p \leq 1\} \subseteq \mathbb{R}^2$, where the first coordinate corresponds to the probability of r and the second coordinate to the probability of s. With abuse of notation, we write $u_i(t_i, (p, 1-p)) = pu_i^r(t_i) + (1-p)u_i^s(t_i)$ for $0 \leq p \leq 1$.

An SCF is a mapping $f : T \to \Delta(A)$. The set of all SCFs is denoted F. It will be useful to think about F as a (convex, compact) subset of the linear space $\mathbb{R}^{2|T|}$. Thus, if $f, g \in F$ and $\alpha \in [0, 1]$ then $\alpha f + (1-\alpha)g \in F$ is defined by $(\alpha f + (1-\alpha)g)(t) = \alpha f(t) + (1-\alpha)g(t) \in \Delta(A)$.

For every agent i, type $t_i \in T_i$ and SCF f we denote by $U_i(f|t_i)$ the interim expected utility of agent i under f conditional on him being of type t_i:

$$U_i(f|t_i) = \mathbb{E}\left(u_i\left(\hat{t}_i, f\left(\hat{t}\right)\right) | \hat{t}_i = t_i\right) = u_i(t_i) \cdot \mathbb{E}\left(f\left(t_i, \hat{t}_{-i}\right)\right),$$

where $x \cdot y$ denotes the inner product of the vectors x and y. The ex ante utility of agent i under f is

$$U_i(f) = \mathbb{E}\left(u_i\left(\hat{t}_i, f\left(\hat{t}\right)\right)\right) = \sum_{t_i \in T_i} \mu_i(t_i) U_i(f|t_i).$$

DEFINITION 1. An SCF f is IC if truth telling is a Bayesian equilibrium of the direct revelation mechanism associated with f. Namely, if for all $i \in N$ and all $t_i, t_i' \in T_i$, we have

$$u_i(t_i) \cdot \left(\mathbb{E}\left(f\left(t_i, \hat{t}_{-i}\right)\right) - \mathbb{E}\left(f\left(t_i', \hat{t}_{-i}\right)\right)\right) \geq 0. \tag{1}$$

The set of all IC SCFs is denoted F^{IC}.

[6] For every finite set X, $\Delta(X)$ denotes the set of probability measures on X.

3.1. *Ordinal Rules and Weighted Majority Rules.* We now define two classes of SCFs that will have an important part in the analysis below. For each agent i, let P_i be the partition of T_i into the two (nonempty) sets

$$T_i^r = \{t_i \in T_i : u_i^r(t_i) > u_i^s(t_i)\},$$
$$T_i^s = \{t_i \in T_i : u_i^r(t_i) < u_i^s(t_i)\}.$$

Recall that agents are never indifferent, so every type t_i is in exactly one of these sets. The partition P_i reflects the ordinal preferences of agent i over the alternatives. Let P be the partition of T, which is the product of all the P_i's: t and t' are in the same element of P if and only if t_i and t_i' are in the same element of P_i for every $i \in N$. As usual, let $P(t)$ be the element of the partition P that contains the type profile t.

DEFINITION 2. An SCF f is ordinal if it is P-measurable, that is, if $f(t) = f(t')$ whenever $P(t) = P(t')$. The set of all ordinal SCFs is denoted F^{ORD}.

Thus, an ordinal SCF depends only on the ordinal information in the reported type profile and is not affected by changes in the expressed intensity of preference.

DEFINITION 3. An SCF f is a weighted majority rule if there are strictly positive numbers $(w_1, \ldots, w_n; q)$ such that $\sum_{i\in N} w_i > q$ and such that

$$f(t) = \begin{cases} (1,0) & \text{if } \sum_{\{i:t_i\in T_i^r\}} w_i > q \\ (0,1) & \text{if } \sum_{\{i:t_i\in T_i^r\}} w_i < q. \end{cases}$$

We refer to w_i as the weight of agent i and to q as the quota.

REMARK 1. Definition 3 does not specify how a weighted majority rule deals with ties, that is, with type profiles for which $\sum_{\{i:t_i\in T_i^r\}} w_i = q$ (when such type profiles exist). In particular, a weighted majority rule need not be ordinal, since it may assign different alternatives to two ordinally equivalent type profiles in which there is a tie. Furthermore, for the very same reason a weighted majority rule need not be IC.

To see this, consider an environment with $n = 2$ agents and where the type of each agent is randomly chosen from the set $\{x, y, z, w\}$ according to a uniform distribution. Both agents have the same utility function: They get utility of 0 if s is chosen no matter what is their type. If r is chosen then their utilities (as a function of their type) are 2 for type x, 1 for type y, -1 for type z, and -2 for type w. Let f be the first-best SCF that chooses the alternative that maximizes the sum of utilities at each type profile and flips a fair coin if there is a tie. First, note that f is a weighted majority rule with weights and quota given by $(1, 1; 1)$. Indeed, the only restriction that this weighted majority rule imposes is that if both agents prefer the same alternative then that alternative is chosen, which is clearly satisfied by f. Second, f is not ordinal since it assigns different outcomes to the ordinally equivalent type profiles (x, z) and (y, w). Third, f is not IC since an agent of type y has an incentive to report that his type is x.

Note, however, that if an SCF is both ordinal and a weighted majority rule then it is IC. Indeed, in a weighted majority rule there is never an incentive to report a type with opposite ranking of alternatives than the actual one, and ordinality implies that an agent cannot gain by pretending to be a different type with the same ordinal preferences.

The following simple lemma provides an alternative representation of weighted majority rules, which will be useful in some of the proofs later on. The proof is straightforward and therefore omitted.

LEMMA 1. *f is a weighted majority rule if and only if there are $2n$ strictly positive numbers $((w_1^r, w_1^s), \ldots, (w_n^r, w_n^s))$ such that*

$$f(t) = \begin{cases} (1,0) & \text{if } \sum_{\{i:t_i\in T_i^r\}} w_i^r > \sum_{\{i:t_i\in T_i^s\}} w_i^s \\ (0,1) & \text{if } \sum_{\{i:t_i\in T_i^r\}} w_i^r < \sum_{\{i:t_i\in T_i^s\}} w_i^s. \end{cases}$$

Moreover, $(w_1, \ldots, w_n; q)$ and $((w_1^r, w_1^s), \ldots, (w_n^r, w_n^s))$ represent the same weighted majority rule whenever $w_i = w_i^r + w_i^s$ for every i and $q = \sum_i w_i^s$.

4. THE UTILITARIAN RULE

We start the analysis by considering the problem of maximizing social welfare, that is, the sum of ex ante expected utilities of all the agents. It will be convenient to denote $v^r(t) = \sum_{i\in N} u_i^r(t_i)$, $v^s(t) = \sum_{i\in N} u_i^s(t_i)$, and $v(t) = (v^r(t), v^s(t))$ for every $t = (t_1, \ldots, t_n) \in T$. These are the welfare totals for each of the two alternatives when t is the realized type profile.

DEFINITION 4. The (ex ante) social welfare of an SCF f is $V(f) = \sum_{i\in N} U_i(f) = \mathbb{E}\left(v\left(\hat{t}\right) \cdot f\left(\hat{t}\right)\right)$.

Without requiring incentive compatibility, a maximizer of social welfare simply chooses r whenever $v^r(t) > v^s(t)$ and chooses s if the other inequality holds (anything, including randomization, can be chosen when $v^r(t) = v^s(t)$). However, such SCFs will typically not be IC, since agents will have an incentive to exaggerate the intensity of their preference. The following theorem characterizes maximizers of social welfare subject to incentive compatibility.

THEOREM 1. *An SCF $f \in F^{IC}$ is a maximizer of V in F^{IC} if and only if it satisfies*

$$f(t) = \begin{cases} (1,0) & \text{if } \sum_{\{i:t_i\in T_i^r\}} \tilde{w}_i^r > \sum_{\{i:t_i\in T_i^s\}} \tilde{w}_i^s \\ (0,1) & \text{if } \sum_{\{i:t_i\in T_i^r\}} \tilde{w}_i^r < \sum_{\{i:t_i\in T_i^s\}} \tilde{w}_i^s, \end{cases} \tag{2}$$

where

$$\tilde{w}_i^r = \mathbb{E}\left(u_i^r(\hat{t}_i) - u_i^s(\hat{t}_i) \mid \hat{t}_i \in T_i^r\right), \; \tilde{w}_i^s = \mathbb{E}\left(u_i^s(\hat{t}_i) - u_i^r(\hat{t}_i) \mid \hat{t}_i \in T_i^s\right). \tag{3}$$

In particular, by Lemma 1, any maximizer of V in F^{IC} is a weighted majority rule with weights and quota given by $\left(\tilde{w}_1^r + \tilde{w}_1^s, \ldots, \tilde{w}_n^r + \tilde{w}_n^s; \sum_i \tilde{w}_i^s\right)$.

According to the theorem, any maximizer of social welfare is a weighted majority rule with specific weights that reflect the *expected* intensity of preference of one alternative over the other. Specifically, given an announced type profile, the agents are partitioned according to whether the type they announced prefers r or s. For each agent who prefers r, the expected utility gain if r is indeed chosen is computed, and this is the weight assigned to this agent. Similarly, agents who prefer s are assigned weights corresponding to their expected gain if indeed s is chosen. The chosen alternative is the one with the larger total weight.

PROOF OF THEOREM 1. The proof is based on the following lemma that shows that, in the class of IC SCFs, restricting attention to ordinal SCFs does not change the interim utility possibility set. A similar result (for the agent-symmetric case) appears in Schmitz and Tröger (2006, Lemma 14).

LEMMA 2. *If* $f \in F^{IC}$ *then*[7] $\mathbb{E}(f|P) \in F^{IC} \cap F^{ORD}$ *and satisfies* $U_i(\mathbb{E}(f|P)|t_i) = U_i(f|t_i)$ *for every* $i \in N$ *and every* $t_i \in T_i$.

The proof of the lemma is in the Appendix. Note that the lemma implies that the same relation holds also in the ex ante stage, that is, $U_i(\mathbb{E}(f|P)) = U_i(f)$ for every $i \in N$. In particular, $V(\mathbb{E}(f|P)) = V(f)$.

We turn now to the proof of the theorem. First, note that for every $g \in F^{ORD}$ we have

$$V(g) = \mathbb{E}\left(v\left(\hat{t}\right) \cdot g\left(\hat{t}\right)\right) = \mathbb{E}\left[\mathbb{E}\left(v\left(\hat{t}\right) \cdot g\left(\hat{t}\right)\middle| P\right)\right] = \mathbb{E}\left[g\left(\hat{t}\right) \cdot \mathbb{E}\left(v\left(\hat{t}\right)\middle| P\right)\right].$$

Thus, an SCF g is a maximizer of V in F^{ORD} if and only if it satisfies

$$g(t) = \begin{cases} (1,0) & \text{if } \mathbb{E}\left(v^r\left(\hat{t}\right) \mid P(t)\right) > \mathbb{E}\left(v^s\left(\hat{t}\right) \mid P(t)\right) \\ (0,1) & \text{if } \mathbb{E}\left(v^r\left(\hat{t}\right) \mid P(t)\right) < \mathbb{E}\left(v^s\left(\hat{t}\right) \mid P(t)\right). \end{cases}$$

Using type independence, this is equivalent to

$$g(t) = \begin{cases} (1,0) & \text{if } \sum_{i\in N} \mathbb{E}\left(u_i^r(\hat{t}_i) \mid P_i(t_i)\right) > \sum_{i\in N} \mathbb{E}\left(u_i^s(\hat{t}_i) \mid P_i(t_i)\right) \\ (0,1) & \text{if } \sum_{i\in N} \mathbb{E}\left(u_i^r(\hat{t}_i) \mid P_i(t_i)\right) < \sum_{i\in N} \mathbb{E}\left(u_i^s(\hat{t}_i) \mid P_i(t_i)\right), \end{cases}$$

which is precisely condition (2) in the theorem.

Now, assume $f \in F^{IC}$ satisfies (2). Then $\mathbb{E}(f|P)$ also satisfies (2), so by the previous argument $\mathbb{E}(f|P)$ is a maximizer of V in F^{ORD}. By Lemma 2, $\mathbb{E}(f|P) \in F^{IC} \cap F^{ORD}$ and $V(f) = V(\mathbb{E}(f|P))$. Assume by contradiction that there is $f' \in F^{IC}$ with $V(f') > V(f)$. Then again by Lemma 2, $\mathbb{E}(f'|P) \in F^{IC} \cap F^{ORD}$ and $V(\mathbb{E}(f'|P)) = V(f')$. It follows that $V(\mathbb{E}(f|P)) = V(f) < V(f') = V(\mathbb{E}(f'|P))$, contradicting the fact that $\mathbb{E}(f|P)$ is a maximizer of V in F^{ORD}.

Conversely, assume that f is a maximizer of V in F^{IC}. Then by Lemma 2, $\mathbb{E}(f|P)$ is also a maximizer of V in F^{IC}. From the first paragraph of this proof, it follows that the maximum of V in F^{ORD} is attained (perhaps not exclusively) by some function in $F^{IC} \cap F^{ORD}$. It follows that $\mathbb{E}(f|P)$ is a maximizer of V in F^{ORD}, and so it must satisfy condition (2). However, if $\mathbb{E}(f|P)$ satisfies (2) then f satisfies (2) as well.

5. PARETO EFFICIENCY

The utilitarian criterion for the evaluation of SCFs is somewhat controversial, as it involves interpersonal comparison of cardinal utilities (see, e.g., the discussion in Weymark, 2005). In contrast, it is hard to argue against Pareto efficiency as a minimal normative criterion that a social planner should take into account. In this section, we will characterize the Pareto frontier of F^{IC} and show that it is closely related to the class of weighted majority rules.

As is well known (see Holmström and Myerson, 1983), in environments with incomplete information the concept of Pareto efficiency is not as simple as in complete information environments. The timing (ex ante, interim, ex post) at which a given SCF is evaluated can affect the conclusion of whether it is efficient or not. Another aspect is whether incentive constraints are taken into account or not. Holmström and Myerson (1983) convincingly argue that, out of the six possible combinations (three possibilities for the timing of evaluation times two possibilities for the set of feasible rules), three are reasonable as notions of efficiency.

DEFINITION 5. (Holmström and Myerson, 1983).

(1) An SCF f is ex ante incentive efficient if $f \in F^{IC}$ and there is no $g \in F^{IC}$ such that $U_i(g) \geq U_i(f)$ for every agent i, with at least one strict inequality.

[7] $\mathbb{E}(f|P)$ is the conditional expectation of f given the partition P, that is, it is the SCF defined by $\mathbb{E}(f|P)(t) = \frac{1}{\mu(P(t))}\sum_{t'\in P(t)} \mu(t')f(t')$.

(2) An SCF f is interim incentive efficient if $f \in F^{IC}$ and there is no $g \in F^{IC}$ such that $U_i(g|t_i) \geq U_i(f|t_i)$ for every agent i and every $t_i \in T_i$, with at least one strict inequality.
(3) An SCF $f \in F$ is ex post classically efficient if there is no $g \in F$ such that $u_i(t_i, g(t)) \geq u_i(t_i, f(t))$ for every agent i and every $t = (t_i, t_{-i}) \in T$, with at least one strict inequality.

The class of ex post classically efficient SCFs is very large. The only restriction that this notion of efficiency imposes in our environment is that of *unanimity*: If a given type profile is such that all the agents prefer one alternative over the other, then this alternative should be chosen. For type profiles in which some agents prefer r and some prefer s any choice (including randomization) is allowed. It follows that this notion of efficiency is too weak to narrow F in a significant way. We therefore focus on the other two criteria. The following theorems characterize interim and ex ante incentive efficiency.

THEOREM 2. *An SCF f is interim incentive efficient if and only if f is an IC weighted majority rule.*[8]

THEOREM 3. *An SCF f is ex ante incentive efficient if and only if there are strictly positive numbers $\lambda_1, \dots, \lambda_n$ such that f is an IC weighted majority rule with weights and quota given by $\left(\lambda_1\left(\tilde{w}_1^r + \tilde{w}_1^s\right), \dots, \lambda_n\left(\tilde{w}_n^r + \tilde{w}_n^s\right); \sum_i \lambda_i \tilde{w}_i^s\right)$, where $((\tilde{w}_1^r, \tilde{w}_1^s), \dots, (\tilde{w}_n^r, \tilde{w}_n^s))$ are the weights used to maximize social welfare as in* (3).

Before the proofs, a couple of comments are in order. First, the weights and quota in an ex ante incentive efficient rule are based on those of the utilitarian rule, but the power of agents may be arbitrarily distorted by the coefficients λ_i. This may be viewed as the result of a social planner maximizing a different welfare functional than the utilitarian. Note, however, that a certain connection between the weights and the quota must hold: If the weight of an agent increases by a certain amount, then the quota should also increase by a certain smaller (environment-specific) amount. Thus, if agent i has greater influence on the outcome when he supports r (his weight increased more than the increase in quota) then he also has greater influence when he supports s (the quota has increased).

One may ask whether, despite the argument of the previous paragraph, interim and ex ante incentive efficiency are in fact equivalent in our simple environment. To see that in general they are not, consider the case where there is symmetry between the alternatives in the sense that $\tilde{w}_i^r = \tilde{w}_i^s$ for every i. In this case, every ex ante incentive efficient rule is *neutral* (except may be in case of a tie); that is, alternative r is chosen when a certain coalition of agents supports it if and only if alternative s is chosen when the same coalition supports s. Obviously, there are many weighted majority rules that do not satisfy this property, and thus the class of interim incentive efficient rules is strictly larger than the class of ex ante incentive efficient rules in this case.

PROOF OF THEOREM 2. Assume first that $f \in F^{IC}$ is a weighted majority rule with corresponding weights $((w_1^r, w_1^s), (w_2^r, w_2^s) \dots, (w_n^r, w_n^s))$ (here we use the alternative representation of a weighted majority rule from Lemma 1). We consider an auxiliary environment with the same set of agents, types, and distribution over types as in the original environment. The utilities in the new environment are given by

$$u_i'(t_i) = \begin{cases} \frac{w_i^r}{\tilde{w}_i^r} u_i(t_i) & \text{if } t_i \in T_i^r \\ \frac{w_i^s}{\tilde{w}_i^s} u_i(t_i) & \text{if } t_i \in T_i^s, \end{cases}$$

[8] We must explicitly require that f be IC; see Remark 1.

where $\tilde{w}_i^r$, $\tilde{w}_i^s$ are as in Equation (3). Since the set of IC SCFs is not affected by this transformation of utilities, we can apply Theorem 1 to conclude that f is a maximizer of V in the auxiliary environment (among IC functions). In other words, f is a maximizer of

$$\sum_i \sum_{t_i} \mu_i(t_i) U_i'(g|t_i)$$

among all functions $g \in F^{IC}$, where $U_i'(g|t_i)$ is the interim utility of agent i of type t_i in the auxiliary environment when g is implemented. Since $U_i'(g|t_i) = \frac{w_i^r}{\tilde{w}_i^r} U_i(g|t_i)$ for $t_i \in T_i^r$ and $U_i'(g|t_i) = \frac{w_i^s}{\tilde{w}_i^s} U_i(g|t_i)$ for $t_i \in T_i^s$, we get that f is a maximizer of

$$\sum_i \left(\sum_{t_i \in T_i^r} \mu_i(t_i) \frac{w_i^r}{\tilde{w}_i^r} U_i(g|t_i) + \sum_{t_i \in T_i^s} \mu_i(t_i) \frac{w_i^s}{\tilde{w}_i^s} U_i(g|t_i) \right).$$

Thus, f maximizes a linear combination with strictly positive coefficients of the interim utilities of the players (in the class F^{IC}). This proves that f is interim incentive efficient.

Conversely, let f be interim incentive efficient. We claim first that there are *strictly positive* numbers $\{\lambda_i(t_i)\}_{i \in N, t_i \in T_i}$ such that f is a maximizer of

$$\sum_i \sum_{t_i} \lambda_i(t_i) U_i(g|t_i)$$

among all functions $g \in F^{IC}$. Indeed, the set F^{IC}, viewed as a subset of $\mathbb{R}^{2|T|}$, is polyhedral and convex. Also, the mapping from SCFs to interim utility vectors is affine: $U_i(\alpha f + (1-\alpha)g|t_i) = \alpha U_i(f|t_i) + (1-\alpha)U_i(g|t_i)$ for any $f, g \in F$ and any $\alpha \in [0,1]$. It follows that the set $\{(U_i(f|t_i))_{i \in N, t_i \in T_i} : f \in F^{IC}\}$ is polyhedral and convex (Rockafellar, 1970, p. 174). For convex polyhedral utility possibility sets, Pareto efficiency is completely characterized by maximization of linear combinations of utilities with strictly positive coefficients (Gale, 1960, p. 308).

Fix a vector λ as above and consider the auxiliary environment with utilities given by $u_i'(t_i) = \frac{\lambda_i(t_i)}{\mu_i(t_i)} u_i(t_i)$ (all other ingredients are the same as in the original environment). As before, incentive compatibility is not affected by this change, and the interim utilities in the new environment are given by $U_i'(g|t_i) = \frac{\lambda_i(t_i)}{\mu_i(t_i)} U_i(g|t_i)$. It follows that f maximizes (in F^{IC}) the expression

$$\sum_i \sum_{t_i} \mu_i(t_i) U_i'(g|t_i) = \sum_i U_i'(g),$$

which means that f is a maximizer of (ex ante) social welfare in the auxiliary environment. By Theorem 1, f must satisfy Equation (2) (for the new environment), which means that f is a weighted majority rule.

PROOF OF THEOREM 3. The proof follows the footsteps of the previous proof. Assume first that $f \in F^{IC}$ is a weighted majority rule with weights and quota as in the theorem (for some vector $(\lambda_1, \ldots, \lambda_n)$ of positive numbers). Consider the auxiliary environment with utilities given by $u_i'(t_i) = \lambda_i u_i(t_i)$ for every i and t_i. It follows from Theorem 1 that f is a welfare maximizer (subject to IC) in the new environment, that is f maximizes the sum

$$\sum_i U_i'(g) = \sum_i \lambda_i U_i(g)$$

over all IC functions g. Since all the λ's are strictly positive we conclude that f is ex ante incentive efficient.

In the other direction, assume that f is ex ante incentive efficient. By an argument similar to the one in the proof of Theorem 2, there are strictly positive numbers $(\lambda_1, \dots, \lambda_n)$ such that f maximizes the sum

$$\sum_i \lambda_i U_i(g)$$

over all IC functions g. It follows that f is socially optimal subject to IC in an auxiliary environment with utilities given by $u_i'(t_i) = \lambda_i u_i(t_i)$. By Theorem 1, f satisfies the condition in the theorem.

6. ANONYMOUS RULES

Voting rules that discriminate between voters, in the sense of giving more power to one group of voters over another, are often excluded as violating the basic fairness criterion of "one person, one vote." In this section, we revisit the questions addressed in the previous sections, under the additional restriction that only SCFs that treat the agents in a symmetric way are feasible. Such SCFs are typically called *anonymous* in the literature, and we will use this terminology as well. We emphasize that we do *not* assume that the environment is symmetric between the agents—different agents may have different utility distributions.

In this section, we restrict attention to ordinal SCFs. This makes the definition of anonymity simpler and allows us to focus on the more interesting aspects of the problem.[9] We denote by $k(t) = \#\{i \in N : t_i \in T_i^r\}$ the number of agents that prefer r over s given the type profile $t = (t_1, \dots, t_n) \in T$.

DEFINITION 6. An SCF f is anonymous if there exists a function $\delta : \{0, 1, \dots, n\} \to \Delta(A)$ such that $f(t) = \delta(k(t))$ for every $t \in T$. The class of anonymous SCFs is denoted F^{ANO}.

The intersection of the classes of weighted majority rules and anonymous rules consists of weighted majority rules in which all agents have the same weight, that is, rules in which the weights and quota are given by $(1, 1, \dots, 1; q)$ for some $0 < q < n$. We refer to such SCFs as *qualified majority rules*. Note that this class contains both super- and submajority rules, so our definition of a qualified majority rule is somewhat different than the usual meaning of this term. Every qualified majority rule f thus has a corresponding function δ of the form

$$\delta(k) = \begin{cases} (0, 1) & \text{if } k < q \\ (1, 0) & \text{if } k > q \end{cases}$$

for some $0 < q < n$. When $k = q$ (in cases where q is integer) every outcome, including randomization, is allowed.

A qualified majority rule is IC since it is an ordinal weighted majority rule (see Remark 1). There are, however, IC anonymous SCFs that are not qualified majority rules, as demonstrated by the following example.

EXAMPLE 1. Let $n = 3$ and consider a symmetric environment in which $T_1 = T_2 = T_3 = \{t_r, t_s\}$, $\mu_i(t_r) = \mu_i(t_s) = 1/2$ for every i, and utilities given by $u_i^r(t_r) = u_i^s(t_s) = 1$ and $u_i^r(t_s) = u_i^s(t_r) = 0$ for every i. Let δ be given by $\delta(0) = \delta(2) = (0, 1)$ and $\delta(1) = \delta(3) = (1, 0)$. That is, if

[9] Alternatively, we could have required that all agents have the same set of type labels and define anonymity by requiring that a permutation of agents' reports does not change the outcome. A modified version of Lemma 2 would then imply that essentially only ordinal SCFs should be considered.

the number of agents that support r is odd then r is chosen, and if it is even then s is chosen. Then (the anonymous SCF generated by) δ is IC even though it is not a qualified majority rule. Indeed, the probability that each of the alternatives is chosen is 1/2 independently of the report of an agent (assuming others are truthful).

Our strategy in this section is similar to the one of the previous sections: We start by characterizing the utilitarian solution within the class of anonymous and IC functions, and we will then use this result to analyze Pareto efficiency in this class. The following theorem generalizes early results in Badger (1972) and Curtis (1972) (see also Lemma 4 in Barberà and Jackson, 2004).

THEOREM 4. *An SCF f is a maximizer of V in the class $F^{IC} \cap F^{ANO}$ if and only if f is a qualified majority rule with a quota $q = \beta\underline{k} + (1-\beta)\overline{k}$ for some $0 < \beta < 1$, where*

$$\overline{k} = \min\left\{k \in \{0, 1, \ldots, n\} : \sum_{i \in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r \geq \sum_{i \in N} \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s\right\}$$

$$\underline{k} = \max\left\{k \in \{0, 1, \ldots, n\} : \sum_{i \in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r \leq \sum_{i \in N} \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s\right\}.$$

The proof of this result is quite simple, since f is a welfare maximizer if and only if for every k, conditional on the event $\{k(t) = k\}$, f chooses the alternative with higher conditional expected welfare. We will show that the difference of the conditional expected welfare between alternative r and alternative s is a strictly increasing function of k, and that $\overline{k}$ ($\underline{k}$) is the minimal (maximal) k at which this function is nonnegative (nonpositive). It follows that an optimal SCF is a qualified majority rule with a quota between these two numbers.[10]

PROOF OF THEOREM 4. We start with the following lemma, which is closely related to Lemma 1 in Barberà and Jackson (2004) and to an earlier work of Badger (1972). The proof can be found in the Appendix.

LEMMA 3. *For each $i \in N$, the conditional probability $\mu\left(T_i^r \mid k(t) = k\right)$ is a strictly increasing function of k.*

Moving to the proof of the theorem, let $f \in F^{ANO}$ and let δ be its associated function. Then

$$V(f) = \sum_{k=0}^{n} \mu\left(k(t) = k\right)\left[\delta(k) \cdot \mathbb{E}\left(v\left(\hat{t}\right) \mid k\left(\hat{t}\right) = k\right)\right].$$

Therefore, f is a maximizer of V if and only if

$$\delta(k) = \begin{cases} (1, 0) & \text{if } \mathbb{E}\left(v^r\left(\hat{t}\right) - v^s\left(\hat{t}\right) \mid k\left(\hat{t}\right) = k\right) > 0 \\ (0, 1) & \text{if } \mathbb{E}\left(v^r\left(\hat{t}\right) - v^s\left(\hat{t}\right) \mid k\left(\hat{t}\right) = k\right) < 0. \end{cases} \tag{4}$$

[10] It is possible that $\overline{k} = \underline{k}$, in which case the quota q is equal to their common value. In this case, an optimal SCF may assign any outcome (including randomization) if exactly q agents prefer r. Otherwise, $\overline{k} = \underline{k} + 1$ and there is a unique deterministic utilitarian rule.

Now,

$$
\begin{aligned}
\mathbb{E}\left(v^r\left(\hat{t}\right) - v^s\left(\hat{t}\right) \mid k\left(\hat{t}\right) = k\right) &= \sum_{i\in N} \mathbb{E}\left(u_i^r\left(\hat{t}_i\right) - u_i^s\left(\hat{t}_i\right) \mid k\left(\hat{t}\right) = k\right) \\
&= \sum_{i\in N} \mu\left(T_i^r \mid k(t) = k\right) \mathbb{E}\left(u_i^r\left(\hat{t}_i\right) - u_i^s\left(\hat{t}_i\right) \mid k\left(\hat{t}\right) = k, \hat{t}_i \in T_i^r\right) \\
&\quad + \sum_{i\in N} \mu\left(T_i^s \mid k(t) = k\right) \mathbb{E}\left(u_i^r\left(\hat{t}_i\right) - u_i^s\left(\hat{t}_i\right) \mid k\left(\hat{t}\right) = k, \hat{t}_i \in T_i^s\right) \\
&= \sum_{i\in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r + \sum_{i\in N} \mu\left(T_i^s \mid k(t) = k\right) (-\tilde{w}_i^s),
\end{aligned}
$$

where in the last equality we used type independence to remove the conditioning event $\{k(t) = k\}$. Thus, (4) becomes

$$
\delta(k) = \begin{cases} (1,0) & \text{if } \sum_{i\in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r > \sum_{i\in N} \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s \\ (0,1) & \text{if } \sum_{i\in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r < \sum_{i\in N} \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s. \end{cases}
$$

Consider the difference

$$
\alpha(k) = \sum_{i\in N} \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r - \sum_{i\in N} \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s.
$$

We have $\alpha(0) = -\sum_{i\in N} \tilde{w}_i^s < 0$ and $\alpha(n) = \sum_{i\in N} \tilde{w}_i^r > 0$. The proof will be complete if we can show that $\alpha(k)$ is strictly increasing in k. A sufficient condition for this is that $\mu\left(T_i^r \mid k(t) = k\right)$ is strictly increasing in k for each i, so we are done by Lemma 3.

We now move on to discuss Pareto efficiency (subject to IC) in F^{ANO}. The definitions of ex ante and interim efficiency are almost the same as those in Definition 5, the only difference being that F^{IC} should be replaced by $F^{IC} \cap F^{ANO}$ everywhere.

THEOREM 5. *An SCF f is interim efficient in $F^{IC} \cap F^{ANO}$ if and only if f is a qualified majority rule.*

THEOREM 6. *An SCF f is ex ante efficient in $F^{IC} \cap F^{ANO}$ if and only if f is a qualified majority rule with a quota $q = \beta \min_{i\in N} \underline{k}_i + (1-\beta) \max_{i\in N} \overline{k}_i$ for some $0 < \beta < 1$, where for each $i \in N$*

$$
\begin{aligned}
\overline{k}_i &= \min\left\{k \in \{0,1,\ldots,n\} : \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r \geq \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s\right\} \\
\underline{k}_i &= \max\left\{k \in \{0,1,\ldots,n\} : \mu\left(T_i^r \mid k(t) = k\right) \tilde{w}_i^r \leq \mu\left(T_i^s \mid k(t) = k\right) \tilde{w}_i^s\right\}.
\end{aligned}
$$

These theorems follow from arguments similar to those of the previous results, so we only give sketches of the proofs. For Theorem 5, notice first that since a qualified majority rule is an IC anonymous weighted majority rule, it follows immediately from Theorem 2 that every qualified majority rule is interim efficient in $F^{IC} \cap F^{ANO}$. Conversely, assume that f is interim efficient in $F^{IC} \cap F^{ANO}$. Then, by an argument identical to the one in the proof of Theorem 2, f maximizes social welfare among all IC functions (not just anonymous) in an auxiliary environment obtained by a monotone transformation of utilities. It follows from Theorem 4 that f is a qualified majority rule.

For Theorem 6, we can use the same argument as in the proof of Theorem 3 to conclude that f is ex ante efficient in $F^{IC} \cap F^{ANO}$ if and only if it maximizes welfare in this set in an environment in which the utility of each agent i is multiplied by some constant $\lambda_i > 0$. Denote by $\underline{k}_\lambda, \overline{k}_\lambda$ the cutoffs of Theorem 4 when agents' utilities are multiplied by $\lambda = (\lambda_1, \dots, \lambda_n)$. Notice that the cutoffs $\underline{k}_i, \overline{k}_i$ in the statement of Theorem 6 are not affected by this transformation of utilities.

Now, for every λ we have that $\min_i \underline{k}_i \le \underline{k}_\lambda \le \overline{k}_\lambda \le \max_i \overline{k}_i$. Thus, if $q = \beta \underline{k}_\lambda + (1-\beta)\overline{k}_\lambda$ for some $\beta \in (0,1)$ then also $q = \beta' \min_i \underline{k}_i + (1-\beta')\max_i \overline{k}_i$ for some $\beta' \in (0,1)$. This proves that every ex ante efficient SCF has the form as in the theorem. Conversely, given $q = \beta' \min_i \underline{k}_i + (1-\beta')\max_i \overline{k}_i$ with $\beta' \in (0,1)$, there exists λ such that $q = \beta \underline{k}_\lambda + (1-\beta)\overline{k}_\lambda$ for some $\beta \in (0,1)$. The latter can be achieved, for example, by assigning appropriate weights to the two agents for which $\min_i \underline{k}_i$ and $\max_i \overline{k}_i$ are achieved and assigning weights close enough to 0 to all other agents.[11]

6.1. *When Are Anonymous Rules Optimal?* We conclude this section by considering the question of what environments have the property that an anonymity constraint does not reduce welfare. That is, in what kind of environments is the optimal rule among all IC rules anonymous?

It is an easy guess that if the environment is symmetric between the agents, in the sense that all agents have the same utility distribution, then the optimal rule is anonymous. The following immediate corollary of Theorem 1 shows that in fact much less symmetry is needed for this result to hold.

COROLLARY 1. *In an environment in which $\tilde{w}_i^r + \tilde{w}_i^s$ is constant in i there is a qualified majority rule that maximizes welfare among all IC SCFs. This rule is characterized by the quota $q = \sum_i \frac{\tilde{w}_i^s}{\tilde{w}_i^r + \tilde{w}_i^s}$.*

The condition in the corollary is far from being necessary. There are very asymmetric environments in which anonymous rules are optimal. Consider, for example, an environment with three agents in which $\tilde{w}_1^r = \tilde{w}_1^s = \tilde{w}_2^r = \tilde{w}_2^s = 100$ and $\tilde{w}_3^r = \tilde{w}_3^s = 1$. Even though agent 3 is very different from agents 1 and 2, the utilitarian rule (in F^{IC}) here is the simple majority rule, so in particular it is anonymous.

We can ask a similar question regarding Pareto efficiency instead of welfare optimality. By Theorem 2, in any environment, every qualified majority rule is interim incentive efficient. What perhaps is less obvious is that in every environment there is at least one ex ante incentive efficient SCF that is anonymous.

COROLLARY 2. *In every environment, the qualified majority rule with quota $q = \sum_i \frac{\tilde{w}_i^s}{\tilde{w}_i^r + \tilde{w}_i^s}$ is ex ante incentive efficient.*

The corollary follows immediately from Theorem 3 by putting $\lambda_i = \frac{1}{\tilde{w}_i^r + \tilde{w}_i^s}$.

7. SELF-STABILITY AND DURABILITY

7.1. *Ex Ante Self-Stability.* Barberà and Jackson (2004) study the self-stability of qualified majority rules. A given status-quo rule f is self-stable if it beats any alternative rule g where the decision between f and g is made using the rule f. Self-stable rules correspond to steady states of the evolutionary process of a constitution. Over time, a rule that is not self-stable is likely to be replaced by another rule, even if the former is excellent from a welfare point of view.

An important question is thus whether any of the weighted majority rules that we study here are self-stable and, if so, which ones. To formally address this question, we need to extend the definition of self-stability to weighted majority rules.

[11] Theorem 6 can also be proved using the "single peakedness" property of the agents' preferences over qualified majority rules; see Badger (1972) and Barberà and Jackson (2004).

DEFINITION 7. A weighted majority rule $f = (w_1, \dots, w_n; q)$ is self-stable if for any other weighted majority rule g it holds that

$$\sum_{i \in R(g,f)} w_i < q,$$

where $R(g,f) = \{i \in N : U_i(g) > U_i(f)\}$ is the set of agents whose ex ante expected utility is higher under g than under f.

We emphasize that this definition is based on an ex ante comparison of the two voting rules. That is, when agents need to decide whether to support the current rule f or the alternative g, they do not know their preferences over the issues that these rules will be used to resolve. Also, notice that implicit in this definition is the assumption that in case of a tie (i.e., when $\sum_{i \in R(g,f)} w_i = q$) the new rule g prevails, but none of our results depend on this tie-breaking rule.

The definition in Barberà and Jackson is the same as above except that both f and g are assumed to be anonymous (i.e., qualified majority rules). Also, they only consider environments in which $\tilde{w}_i^r = \tilde{w}_i^s = 1$ for all i. They prove (Theorem 1) that if $\mu\left(T_i^r\right)$ is the same for all i then simple majority[12] is self-stable. In contrast, when weighted majority rules are allowed we get the following result.

PROPOSITION 1. *Consider an environment in which $\tilde{w}_i^r = \tilde{w}_i^s = 1$ for all i, and such that $\mu\left(T_i^r\right) = p \in (0,1)$ for all i. If the number of agents is $n \geq 4$ then there is no self-stable qualified majority rule. In particular, the utilitarian rule in such environments is not self-stable.*[13]

PROOF. First, any qualified majority rule different from simple majority is not self-stable since it is defeated by simple majority: It follows from our Theorem 1 that simple majority is the utilitarian rule in this environment and that it gives strictly higher welfare than any other qualified majority rule. From the symmetry between the agents, it follows that every agent gets strictly higher expected utility under simple majority than under any other qualified majority rule. Thus, all the agents will support a shift from a status-quo qualified majority rule to simple majority, and so the qualified majority is not self-stable.

Thus, to complete the proof we only need to check that simple majority is not self-stable. For this, we need the following lemma, whose proof is in the Appendix.

LEMMA 4. *Let $X_1, \dots, X_n$ ($n \geq 4$) be independent and identically distributed random variables with $Pr(X_1 = 1) = 1 - Pr(X_1 = 0) = p \in (0,1)$, and for each $1 \leq k \leq n$ denote $S_k = \sum_{i=1}^k X_i$. Then*

$$\Pr\left(X_1 = 1, S_n \geq \frac{n}{2}\right) + \Pr\left(X_1 = 0, S_n < \frac{n}{2}\right) < \Pr\left(X_1 = 1, S_{n-2} \geq \frac{n-2}{2}\right) + \Pr\left(X_1 = 0, S_{n-2} < \frac{n-2}{2}\right). \tag{5}$$

The left-hand side of (5) is the ex ante expected utility of agent 1 under simple majority rule when there are n agents and the tie-breaking rule favors alternative r (when n is even). Consider the weighted majority rule in which agents 1 through $n-2$ each has a weight of 1, agents $n-1$

[12] We call simple majority to the qualified majority rule in which the quota is half of the total number of agents n. If n is odd then there is a unique simple majority rule, whereas if n is even there are two such (deterministic) rules, depending on how a tie is broken. We refer to both of them as simple majority.

[13] With $n = 3$ agents, it is easy to check that simple majority is self-stable. With $n = 2$ agents, the qualified majority rule that requires unanimous support for a reform is self-stable.

and n each has a weight of $\epsilon < 1/4$, the quota is $(n-2)/2$, and the tie-breaking rule still favors r. The ex ante expected utility of agent 1 is then given by the right-hand side of (5). It follows from the lemma that agent 1 prefers the latter weighted majority rule over simple majority with the tie-breaking rule that favors r. By symmetry, the same is true for agents $2, \ldots, n-2$. This proves that simple majority in which the tie-breaking rule favors r is not self-stable. But notice that the tie-breaking rule of simple majority does not affect the ex ante utility of agents in our environment, so the same is true if the tie-breaking rule favors s. This completes the proof. ■

The proof of the proposition makes it clear that self-stability becomes a very restrictive notion once weighted majority rules can be used to challenge the status-quo rule. Keeping the assumptions of Proposition 1, even in the larger class of nonanonymous rules only a few are self-stable. In other environments there may be more self-stable rules. A precise characterization of self-stability is an interesting direction for future research.

7.2. *Durability and Interim Self-Stability.* A different notion of stability of SCFs, called *durability*, was introduced by Holmström and Myerson (1983). Roughly speaking, an SCF f is durable if f would never (at any information state) be unanimously rejected in favor of another SCF g. As opposed to self-stability of the previous subsection, the decision between the status-quo f and the alternative g takes place in the interim stage, after each agent knows his type. This is a crucial difference, since when an agent decides between the two rules he needs to take into account not only his own information, but also the fact that other agents' voting behavior depends on their information. Indeed, learning that the alternative rule g defeated f (or that the complementary event happened) contains information about other agents' types and, therefore, about how they are going to behave in each of these mechanisms.

The precise definition of durability is quite involved, and we will not give the details here. Basically, given the current rule $f \in F^{IC}$ and some alternative $g \in F$, the play of the direct mechanism f is augmented by a preliminary voting game in which each agent can vote for either f or g, and g replaces f only if g gets a unanimous support. f is called durable if, for every alternative g, there is a sequential equilibrium (Kreps and Wilson, 1982) of the augmented game in which g is never chosen. The following is a corollary of Theorem 2.

COROLLARY 3. *Every ordinal weighted majority rule is durable.*

PROOF. It is proved in Holmström and Myerson (1983, Theorem 2) that any interim incentive efficient and strategy-proof SCF is durable. Strategy-proofness means that truth telling is not just a Bayesian equilibrium of the direct revelation mechanism, it is also a dominant strategy for each agent. Formally,

$$u_i(t_i) \cdot (f(t_i, t_{-i}) - f(t_i', t_{-i})) \geq 0$$

for all i, t_i, t_i', and t_{-i}. It is straightforward to check that any ordinal weighted majority rule is strategy-proof. By Theorem 2, any such SCF is interim incentive efficient, so the proof is complete. ■

A natural modification of durability in our framework would be *interim self-stability*, which replaces the unanimity rule in the preliminary voting stage by the current status-quo rule f. That is, if f is the weighted majority rule $f = (w_1, \ldots, w_n; q)$, then g defeats f if and only if $\sum w_i \geq q$, where the sum is taken over all agents who voted for g. This modification makes it easier for g to beat the incumbent rule f, so interim self-stability is a more demanding concept than durability. An analysis of this stability concept is beyond the scope of this article and we leave it for future research.

8. APPLICATIONS

8.1. *Representatives of Groups with Heterogenous Sizes.* The design of voting rules in a representative democracy with heterogenous country sizes has been thoroughly analyzed in Barberà and Jackson (2006). The variance in country size leads very naturally to an asymmetric environment at the representatives level. Therefore, asymmetric voting rules are appealing from a theoretical point of view in such institutions, and there are several examples of such asymmetric rules being used (e.g., the Council of the European Union). Here, we explain the relation between the model and results in Barberà and Jackson (2006) and those of this article and discuss some of the implications of our results for such environments.

In Barberà and Jackson (2006), a representative of each country votes for one of two alternatives. Her vote is a function of the realized (cardinal) preference profile of the citizens in the country she represents. The voting rule aggregates the votes of the representatives and chooses one of the alternatives (possibly at random). Agents in our model thus correspond to the representatives, and the type of an agent corresponds to a realization of a preference profile in that representative's country.

A main difference between the models is that Barberà and Jackson restrict the message space of agents in the mechanism to consist of only two messages—each agent can only announce the alternative she supports. They find the optimal (utilitarian) rule for this class of mechanisms. Our approach is to consider the direct mechanism in which agents simply report their types. The advantage of this latter approach is that, by the revelation principle, there is no mechanism that can implement an SCF that yields a higher welfare than in the optimal solution of the direct mechanism. Therefore, our Theorem 1 strengthen the conclusion of Theorem 1 in Barberà and Jackson (2006) by showing that the weighted majority rule they find is optimal not just in the class of mechanisms they consider but universally among all possible mechanisms.[14]

Our Theorems 2 and 3 provide further insight regarding the class of rules that should be considered by a social planner in such environments. Theorem 2 implies that only weighted majority rules should be considered, whereas Theorem 3 gives some guidance as to what weights and quota to choose. For concreteness, let us consider the "common bias" model analyzed in Barberà and Jackson (2006). This is the special case in which $\tilde{w}_i^s = \gamma \tilde{w}_i^r$ for every $i \in N$, where $\gamma > 0$ is a parameter. The utilitarian rule in this case is a weighted majority rule with weights and quota given by $(\tilde{w}_1^r, \ldots, \tilde{w}_n^r; \frac{\gamma}{1+\gamma} \sum_i \tilde{w}_i^r)$. It follows from Theorem 3 that ex ante incentive efficiency is exactly characterized by weighted majority rules in which the quota is a $\frac{\gamma}{1+\gamma}$ fraction of the total weight. Thus, although there may be disagreement about whether the utilitarian rule with its specific weights should be used, it seems quite reasonable to focus on rules in which the quota is as implied by Pareto efficiency considerations.

Finally, assume that despite the size heterogeneity, only anonymous rules can be used. This is the practice in the General Assembly of the United Nations (UN), for example. Under the above common bias assumption, it follows from Theorem 4 that the utilitarian rule is then a qualified majority rule with a quota anywhere between

$$\overline{k} = \min \left\{ k : \sum_{i \in N} \mu \left(T_i^r \mid k(t) = k \right) \tilde{w}_i^r \geq \frac{\gamma}{1+\gamma} \sum_{i \in N} \tilde{w}_i^r \right\}$$

and

$$\underline{k} = \max \left\{ k : \sum_{i \in N} \mu \left(T_i^r \mid k(t) = k \right) \tilde{w}_i^r \leq \frac{\gamma}{1+\gamma} \sum_{i \in N} \tilde{w}_i^r \right\}.$$

[14] For the analogy between the models to be accurate, we need to assume that the preferences of the representatives coincide with the aggregate preference in their countries. Barberà and Jackson are not concerned with incentive compatibility of their rule, and they allow representatives to vote arbitrarily. They do, however, impose that representatives vote in accordance with their country's preference when they analyze the more specific "block model."

Note that the probabilities $\mu(T_i^r)$ that country i prefers the reform, who do not play any role when nonanonymous rules are considered, become important to determine the optimal quota. If we add the assumption that $\mu(T_i^r)$ does not depend on i, then $\mu(T_i^r \mid k(t) = k) = k/n$ and the optimal quota becomes $\frac{\gamma}{1+\gamma}n$. In the General Assembly of the UN, passing a resolution on some of the more significant issues requires a 2/3 majority, which is optimal (under our assumptions) if $\gamma = 2$. For other issues, a simple majority is used, which is optimal for $\gamma = 1$.

8.2. *Dichotomous Populations.* Consider, for the sake of concreteness, an academic department that needs to decide whether to give a job offer to a certain faculty candidate (alternative r) or not (alternative s). The current faculty of the department (the voters) has members that work in the same field as the candidate (group G_1) and members that work in other fields (group G_2). Assume that members of G_1 have more to gain (or lose) from the decision than members of G_2 in the sense that there are two numbers $h > l > 0$ such that $\tilde{w}_i^r = \tilde{w}_i^s = h$ for $i \in G_1$ and $\tilde{w}_i^r = \tilde{w}_i^s = l$ for $i \in G_2$.

The utilitarian rule here is to assign a weight of h to each member of G_1, a weight of l to each member of G_2, and set the quota to be half of the total weight, that is, $q = \frac{|G_1|h+|G_2|l}{2}$. Note that if h and l are sufficiently close then the optimal rule is a simple majority rule, whereas if h is much larger than l then the candidate's own field members make the decision by themselves (by a simple majority).

Consider now the more realistic case where only anonymous rules can be used for the decision. As above, the probabilities $\mu(T_i^r)$ become relevant for determining the optimal quota. Assume that these probabilities are symmetric within each group such that there are $0 < p_l \leq p_h < 1$ with $\mu(T_i^r) = p_h$ for $i \in G_1$ and $\mu(T_i^r) = p_l$ for $i \in G_2$. For $i \in G_1$ denote $p_h(k) = \mu(T_i^r \mid k(t) = k)$ and for $i \in G_2$ denote $p_l(k) = \mu(T_i^r \mid k(t) = k)$. Then by Theorem 4, the optimal quota is, roughly speaking, the number k for which $|G_1|hp_h(k) + |G_2|lp_l(k)$ is closest to $\frac{|G_1|h+|G_2|l}{2}$.

Solving for the optimal quota explicitly is very complicated, but we can say something in the two extreme cases when (1) p_h is close to p_l and (2) p_h is close to 1 and p_l is close to 0. In the first case, we are back in the situation analyzed in the previous subsection with $\gamma = 1$, so the optimal anonymous rule is a simple majority. In the second case, we can approximate $p_h(k)$ and $p_l(k)$ as follows: If $k \leq |G_1|$ then conditional on $k(t) = k$ it is almost certain that the entire group of k supporters comes from G_1. It follows that $p_h(k) \approx \frac{k}{|G_1|}$ and $p_l(k) \approx 0$. If $k > |G_1|$ then conditional on $k(t) = k$ it is almost certain that the group of k supporters is composed of the G_1 group plus $k - |G_1|$ members of G_2, so $p_h(k) \approx 1$ and $p_l(k) \approx \frac{k-|G_1|}{|G_2|}$. The optimal quota then depends on which of the two groups has a larger total expected preference intensity, that is, which of the two numbers $|G_1|h$ or $|G_2|l$ is bigger: If $|G_1|h > |G_2|l$ then the optimal quota is (roughly) $q = \frac{|G_1|}{2} + \frac{|G_2|l}{2h} \leq |G_1|$, so if the field members unanimously agree to hire the candidate then a job offer should be given, even if all members of $|G_2|$ object. If, on the other hand, $|G_1|h < |G_2|l$ then the optimal quota is $q = |G_1| + \frac{|G_2|}{2} - \frac{|G_1|h}{2l} > |G_1|$, and at least some support from G_2 is required to pass the threshold.

8.3. *Veto Players.* There are 15 members of the UN Security Council, of which 5 are permanent members and 10 nonpermanent members that are elected to serve two-year terms. To approve any substantial proposal, the affirmative votes of at least 9 of the 15 members are required. In addition, if any 1 of the 5 permanent members casts a negative vote (veto), then the proposal is not adopted. This voting rule is in fact a weighted majority rule:[15] Denote by P the set of permanent members and by NP the set of nonpermanent members. Set $w_i = 100$ for $i \in P$, $w_i = 2$ for $i \in NP$, and $q = 507$. Thus, in order to pass, a proposal needs the support of the 5 permanent members and at least 4 of the nonpermanent members.

[15] Members of the Security Council can also cast an "abstain" vote. We ignore here this possibility since it complicates the analysis. Note, however, that, by the revelation principle, any SCF that can be implemented by a mechanism that allows for abstention can also be implemented by the direct mechanism we consider.

In what kind of environments is this weighted majority rule welfare maximizing? To apply Theorem 1, we need to represent this rule as in Lemma 1. This can be done, for example, by defining $w_i^r = 1$, $w_i^s = 99$ for $i \in P$, and $w_i^r = 0.8$, $w_i^s = 1.2$ for $i \in NP$. By Theorem 1, in environments in which $\tilde{w}_i^r = w_i^r$ and $\tilde{w}_i^s = w_i^s$ for each i, where w_i^r, w_i^s have the above values, the voting rule used in the Security Council is optimal.

Of course this representation of the voting rule is not unique—there are infinitely many different representations. But in any representation it must be the case that for the permanent members w_i^r is much smaller than w_i^s. Thus, for this voting rule to be optimal the environment must have the property that for any permanent member the expected utility gain from adopting a proposal conditional on supporting the proposal is much smaller than the expected utility loss from adopting a proposal conditional on objecting to it.

9. DISCUSSION

9.1. *More Than Two Alternatives.* In this article, we restricted attention to the case of a binary alternative set. With three or more alternatives, the analysis becomes significantly more complicated. With two alternatives, Lemma 2 tells us that we can essentially consider only ordinal SCFs. When there are three or more alternatives, Lemma 2 is no longer true: The utility possibility set generated by IC ordinal functions is a strict subset of that generated by all IC functions and, furthermore, the Pareto frontier of these two sets may be different. Below we give a simple example to illustrate this point, and we refer the reader to Kim (2012) for more general results on the possibility to use cardinal information to improve welfare when there are more than two alternatives.[16]

Let $N = \{1, 2\}$, $T_1 = T_2 = \{x, y, z, w\}$, and assume that the type of each agent i is uniformly (and independently) drawn from T_i. There are three alternatives $A = \{a, b, c\}$. For $i = 1, 2$, utilities are given by $u_i(x) = (u_i(x, a), u_i(x, b), u_i(x, c)) = (5, 4, 1)$, $u_i(y) = (5, 2, 1)$, $u_i(z) = (-5, -4, -1)$, and $u_i(w) = (-5, -2, -1)$. Notice that types x, y and types z, w have the same ordinal preferences over alternatives.

Let f^* be the ex post first-best SCF, that is, the function that chooses an alternative that maximizes the sum of utilities at every type profile, and assume that in case of a tie the winning alternative is randomly chosen from the set of maximizers. It is tedious but straightforward to check that f^* is IC. Thus, f^* is the utilitarian solution and, in particular, it is both ex ante and interim incentive efficient.

However, we claim that there is no IC ordinal SCF f such that $U_i(f|t_i) = U_i(f^*|t_i)$ for every t_i and every i. Indeed, if f is such a function then it must also be a maximizer of ex ante social welfare. Thus, f must choose an alternative that maximizes social welfare at every type profile. However, in type profile (x, w) the unique maximizer is alternative b, whereas in type profile (y, z) the maximizers are alternatives a and c. If f is ordinal then it must choose the same alternative in these two type profiles, a contradiction.

9.2. *Correlated Types.* The following example shows the importance of the type-independence assumption for our results.[17] Consider a three agents environment with $T_1 = T_2 = T_3 = \{x, y, z, w\}$, and utilities given by $u_i(x) = (3, 0)$, $u_i(y) = (1, 0)$, $u_i(z) = (0, 1)$, and $u_i(w) = (0, 3)$ for $i = 1, 2, 3$. The distribution of type profiles is given by $\mu(x, z, z) = \mu(z, x, z) = \mu(z, z, x) = \mu(w, y, y) = \mu(y, w, y) = \mu(y, y, w) = \frac{1}{6}$.

It will be convenient to define a metric d over the set of type profiles T by letting $d(t, t') = \#\{1 \le i \le 3 : t_i \ne t_i'\}$. Denote $E_1 = \{(x, z, z), (z, x, z), (z, z, x)\}$ and $E_2 = \{(w, y, y), (y, w, y), (y, y, w)\}$. Define an SCF f^* as follows: If $t \in E_1$ or $d(t, t') = 1$ for some

[16] The example below shares some similarities with the problem analyzed in Börgers and Postl (2009).

[17] For a different example demonstrating the effect of type correlation on the optimal SCF, see Example 3 in Schmitz and Tröger (2011).

$t' \in E_1$ then $f^*(t) = r$. If $t \in E_2$ or $d(t, t') = 1$ for some $t' \in E_2$ then $f^*(t) = s$. Otherwise, $f^*(t)$ is defined in an arbitrary way. Notice that f^* is well defined since $d(t, t') = 3$ for every $t \in E_1$ and $t' \in E_2$.

First, f^* is IC since no player can influence the outcome by reporting untruthfully (assuming other agents are truthful). Second, f^* coincides with the ex post first-best SCF on $E_1 \cup E_2$. Since $\mu(E_1 \cup E_2) = 1$ this implies that f^* is a maximizer of social welfare (and so incentive Pareto efficient). However, it is not hard to see that f^* is not a weighted majority rule. Indeed, assume by contradiction that $(w_1, w_2, w_3; q)$ represent f^*. Since $f^*(x, z, z) = r$ we must have $w_1 \geq q$, but since $f^*(y, y, w) = s$ we must have $w_1 + w_2 \leq q$. Since all weights must be strictly positive, these two inequalities cannot hold simultaneously.

The above example relies on the fact that most type profiles can never occur. It is thus quite fragile. But it is possible to modify this example to get a similar result with a distribution μ that has a full support. Indeed, fix two small numbers $\delta, \epsilon > 0$. Assume that the probability of each type profile in $E_1 \cup E_2$ is $\frac{1-\epsilon}{6}$, and that the rest of the probability (ϵ) is evenly distributed among the rest of the 58 type profiles. For $t \in E_1$, define $f(t) = (1 - \delta, \delta)$, and for $t \in E_2$ let $f(t) = (\delta, 1 - \delta)$. If t contains at least two agents of type z (but $t \notin E_1$) then $f(t) = (1 - 2\delta, 2\delta)$, and if t contains at least two agents of type y (but $t \notin E_2$) then $f(t) = (2\delta, 1 - 2\delta)$. If t contains a type x agent and a type z agent (but $t \notin E_1$) then $f(t) = (1, 0)$, and if t contains a type y agent and a type w agent (but $t \notin E_2$) then $f(t) = (0, 1)$. Finally, f is defined in an arbitrary way for other type profiles. It is not hard to check that if δ and ϵ are sufficiently small then f yields higher expected welfare than any ordinal rule (in particular higher than any weighted majority rule). It is tedious but straightforward to check that f is IC when ϵ is much smaller than δ.

9.3. *Weighted Majority Rules with Zero Weights.* In the definition of a weighted majority rule (Definition 3), we require that the weights of all players be strictly positive. It turns out that this is essential for the result to hold: If some of the players have zero weights then the SCF may not be incentive interim efficient. The reason is that we may have more type profiles in which there is exactly a tie, and the tie-breaking rule may introduce inefficiencies.

To see this, consider an environment with six agents, each of which is equally likely to be of type t_r or type t_s. Utilities are given by $u_i(t_r) = (1, 0)$ and $u_i(t_s) = (0, 1)$, so t_r prefers r whereas t_s prefers s. Consider the weighted majority rule f given by the weights and quota $(1, 1, 1, 1, 0, 0; 2)$. Thus, r is chosen if at least three agents out of the group $\{1, 2, 3, 4\}$ prefer r, and s is chosen if at most one agent out of $\{1, 2, 3, 4\}$ prefers r.

To complete the description of the rule f, we need to specify which alternative is chosen if exactly two agents out of $\{1, 2, 3, 4\}$ prefer r. Let us introduce the following tie-breaking rule: If the coalitions that prefer r are $\{1, 2\}$, $\{1, 3\}$, or $\{2, 4\}$ then r is chosen; otherwise s is chosen. Notice that this rule ignores completely the preferences of agents 5 and 6.

We claim that we can introduce a different tie-breaking rule such that the resulting SCF g interim Pareto dominates f. Indeed, let the minimal winning coalitions in g (in case of a tie) be $\{1, 2\}, \{1, 3, 5\}, \{1, 3, 6\}, \{1, 4, 5, 6\}, \{2, 4, 5\}, \{2, 4, 6\}$, and $\{2, 3, 5, 6\}$. The change from f to g occurs in four type profiles: $(t_r, t_s, t_r, t_s, t_s, t_s)$, $(t_s, t_r, t_s, t_r, t_s, t_s)$, $(t_r, t_s, t_s, t_r, t_r, t_r)$, and $(t_s, t_r, t_r, t_s, t_r, t_r)$. In the first two, the outcome is switched from r to s, whereas in the last two from s to r. Because of the symmetry of the problem, it is easy to see that the interim utilities of agents 1, 2, 3, and 4 are not affected by this change. However, since the new tie-breaking rule is responsive to the preferences of agents 5 and 6, their interim utilities increase.

9.4. *Implementation.* Reporting truthfully is an equilibrium of the direct revelation mechanism associated with a weighted majority rule, but there are also other equilibria and, furthermore, in some of these equilibria the chosen alternative may be different than under truthfulness. In fact, in many cases there will be no mechanism that implements a given weighted majority rule as the unique equilibrium. The problem is that in many cases a weighted majority rule will fail to satisfy Bayesian monotonicity, which is a necessary condition for implementation. See the closely related Example 2 in Palfrey and Srivastava (1989) for details.

Note, however, that if one strengthens the solution concept to require players to play a Bayesian equilibrium using strategies that are *not weakly dominated* then the problem of equilibrium multiplicity disappears (here we assume that the tie-breaking rule is ordinal). Furthermore, truth telling is a *weakly dominant* strategy, which makes this equilibrium particularly plausible. See Jackson (1991) and Palfrey and Srivastava (1989) for more on implementation in environments with incomplete information.

APPENDIX

A.1. *Proof of Lemma 2.* Let $f \in F^{IC}$. From (1), we have that

$$u_i(t_i) \cdot \left(\mathbb{E}\left(f(t_i, \hat{t}_{-i})\right) - \mathbb{E}\left(f(t_i', \hat{t}_{-i})\right)\right) \geq 0$$

and

$$u_i(t_i') \cdot \left(\mathbb{E}\left(f(t_i, \hat{t}_{-i})\right) - \mathbb{E}\left(f(t_i', \hat{t}_{-i})\right)\right) \leq 0$$

for every i and every $t_i, t_i' \in T_i$. Denote $(p, 1-p) = \mathbb{E}\left(f(t_i, \hat{t}_{-i})\right)$ and $(q, 1-q) = \mathbb{E}\left(f(t_i', \hat{t}_{-i})\right)$. Then the above inequalities become

$$(p-q)(u_i^r(t_i) - u_i^s(t_i)) \geq 0$$

and

$$(q-p)(u_i^r(t_i') - u_i^s(t_i')) \geq 0,$$

respectively. Now, if both t_i, t_i' are in T_i^r or both are in T_i^s then the above inequalities imply that $p = q$, that is $\mathbb{E}(f(t_i, \hat{t}_{-i})) = \mathbb{E}(f(t_i', \hat{t}_{-i}))$. In other words, $\mathbb{E}(f(t_i, \hat{t}_{-i}))$ viewed as a function of t_i is constant on T_i^r and on T_i^s (P_i-measurable).

Denote $g = \mathbb{E}(f|P)$. Obviously, $g \in F^{ORD}$. For any $t_i \in T_i^r$, we have

$$\mathbb{E}\left(f(t_i, \hat{t}_{-i})\right) = \mathbb{E}\left(f\left(\hat{t}\right) | \hat{t}_i \in T_i^r\right) = \mathbb{E}\left(g\left(\hat{t}\right) | \hat{t}_i \in T_i^r\right) = \mathbb{E}\left(g(t_i, \hat{t}_{-i})\right), \tag{A.1}$$

where the first equality follows from the fact that $\mathbb{E}(f(t_i, \hat{t}_{-i}))$ is constant on T_i^r and type independence, the second from the definition of g and the fact that $\{\hat{t}_i \in T_i^r\}$ is in (the algebra generated by) P, and the third from the fact that $g(\cdot, t_{-i})$ is constant on T_i^r for any fixed t_{-i} and type independence. Using the same argument for $t_i \in T_i^s$, we get that (A.1) is valid for every $t_i \in T_i$. Therefore,

$$U_i(f|t_i) = u_i(t_i) \cdot \mathbb{E}\left(f(t_i, \hat{t}_{-i})\right) = u_i(t_i) \cdot \mathbb{E}\left(g(t_i, \hat{t}_{-i})\right) = U_i(g|t_i),$$

which establishes that every type of every agent gets the same interim utility under f and under g. Finally, the fact that g is IC immediately follows from (1) and (A.1).

A.2. *Proof of Lemma 3.* Fix $i \in N$ and denote $k_{-i}(t) = \#\{j \in N \setminus \{i\} : t_j \in T_j^r\}$ the number of agents other than i that prefer alternative r. Then

$$\mu\left(T_i^r \mid k(t) = k\right) = \frac{\mu\left(T_i^r\right)\mu\left(k_{-i}(t) = k-1\right)}{\mu\left(T_i^r\right)\mu\left(k_{-i}(t) = k-1\right) + \mu\left(T_i^s\right)\mu\left(k_{-i}(t) = k\right)} = \frac{1}{1 + \frac{\mu\left(T_i^s\right)\mu(k_{-i}(t)=k)}{\mu\left(T_i^r\right)\mu(k_{-i}(t)=k-1)}}.$$

It follows immediately from Inequality (6) in Samuels (1965, p. 1272) that this ratio is strictly increasing in k, so the proof is complete.

A.3. *Proof of Lemma 4.* We have

$$\begin{aligned}
&\Pr\left(X_1 = 1, S_n \geq \frac{n}{2}\right) - \Pr\left(X_1 = 1, S_{n-1} \geq \frac{n-1}{2}\right)\\
&= p\left[\Pr\left(S_{n-1} \geq \frac{n-2}{2}\right) - \Pr\left(S_{n-2} \geq \frac{n-3}{2}\right)\right]\\
&= p\left[\Pr(X_{n-1} = 1)\Pr\left(S_{n-2} \geq \frac{n-4}{2}\right) + \Pr(X_{n-1} = 0)\Pr\left(S_{n-2} \geq \frac{n-2}{2}\right) - \Pr\left(S_{n-2} \geq \frac{n-3}{2}\right)\right]\\
&= p^2\left[\Pr\left(S_{n-2} \geq \frac{n-4}{2}\right) - \Pr\left(S_{n-2} \geq \frac{n-2}{2}\right)\right] + p\left[\Pr\left(S_{n-2} \geq \frac{n-2}{2}\right) - \Pr\left(S_{n-2} \geq \frac{n-3}{2}\right)\right]\\
&= p^2\Pr\left(\frac{n-4}{2} \leq S_{n-2} < \frac{n-2}{2}\right) - p\Pr\left(\frac{n-3}{2} \leq S_{n-2} < \frac{n-2}{2}\right).
\end{aligned}$$

A similar computation gives

$$\begin{aligned}
&\Pr\left(X_1 = 0, S_n < \frac{n}{2}\right) - \Pr\left(X_1 = 0, S_{n-1} < \frac{n-1}{2}\right)\\
&\quad = -p(1-p)\Pr\left(\frac{n-2}{2} \leq S_{n-2} < \frac{n}{2}\right) + (1-p)\Pr\left(\frac{n-1}{2} \leq S_{n-2} < \frac{n}{2}\right).
\end{aligned}$$

Assume first that n is even. Then the sum of these two differences becomes

$$\begin{aligned}
&p^2 \Pr\left(S_{n-2} = \frac{n-4}{2}\right) - p(1-p)\Pr\left(S_{n-2} = \frac{n-2}{2}\right)\\
&\quad = \binom{n-2}{\frac{n-4}{2}} p^{\frac{n}{2}}(1-p)^{\frac{n}{2}} - \binom{n-2}{\frac{n-2}{2}} p^{\frac{n}{2}}(1-p)^{\frac{n}{2}} < 0.
\end{aligned}$$

If n is odd then after some simple manipulations the sum of the two differences becomes

$$\begin{aligned}
&(1-p)^2 \Pr\left(S_{n-2} = \frac{n-1}{2}\right) - p(1-p)\Pr\left(S_{n-2} = \frac{n-3}{2}\right)\\
&\quad = \binom{n-2}{\frac{n-1}{2}} p^{\frac{n-1}{2}}(1-p)^{\frac{n+1}{2}} - \binom{n-2}{\frac{n-3}{2}} p^{\frac{n-1}{2}}(1-p)^{\frac{n+1}{2}} = 0.
\end{aligned}$$

To conclude, we have shown that

$$\begin{aligned}
&\Pr\left(X_1 = 1, S_n \geq \frac{n}{2}\right) + \Pr\left(X_1 = 0, S_n < \frac{n}{2}\right)\\
&\leq \Pr\left(X_1 = 1, S_{n-1} \geq \frac{n-1}{2}\right) + \Pr\left(X_1 = 0, S_{n-1} < \frac{n-1}{2}\right),
\end{aligned}$$

with a strict inequality when n is even and equality when n is odd. Applying twice this inequality, we get (5).

REFERENCES

APESTEGUIA, J., M. A. BALLESTER, AND R. FERRER, “On the Justice of Decision Rules,” *Review of Economic Studies* 78 (2011), 1–16.

BADGER, W. W., “Political Individualism, Positional Preferences, and Optimal Decision Rules,” in R. G. Niemi and H. F. Weisberg, eds., *Probability Models of Collective Decision Making* (Columbus, Ohio: Merrill, 1972), 34–59.

BARBERÀ, S., AND M. O. JACKSON, “Choosing How to Choose: Self-Stable Majority Rules and Constitutions,” *Quarterly Journal of Economics* 119 (2004), 1011–48.

———, AND ———, “On the Weights of Nations: Assigning Voting Weights in a Heterogeneous Union,” *The Journal of Political Economy* 114 (2006), 317–39.

BEISBART, C., AND L. BOVENS, “Welfarist Evaluations of Decision Rules for Boards of Representatives,” *Social Choice and Welfare* 29 (2007), 581–608.

BÖRGERS, T., AND P. POSTL, “Efficient Compromising,” *Journal of Economic Theory* 144 (2009), 2057–76.

CURTIS, R. B., “Decision Rules and Collective Values in Constitutional Choice,” in R. G. Niemi and H. F. Weisberg, eds., *Probability Models of Collective Decision Making* (Columbus, Ohio: Merrill, 1972), 23–33.

FLEURBAEY, M., “One Stake One Vote,” Mimeo, University Paris-Descartes, 2008.

GALE, D., *The Theory of Linear Economic Models* (New York: McGraw-Hill, 1960).

GERSHKOV, A., B. MOLDOVANU, AND X. SHI, “Optimal Mechanism Design without Money,” Mimeo, University of Toronto, 2013.

HOLMSTRÖM, B., AND R. B. MYERSON, “Efficient and Durable Decision Rules with Incomplete Information,” *Econometrica* 51 (1983), 1799–819.

JACKSON, M. O., “Bayesian Implementation,” *Econometrica* 59 (1991), 461–77.

———, AND H. F. SONNENSCHEIN, “Overcoming Incentive Constraints by Linking Decisions,” *Econometrica* 75 (2007), 241–57.

KIM, S., “Ordinal versus Cardinal Voting Rules: A Mechanism Design Approach,” Mimeo, The Ohio State University, 2012.

KREPS, D. M., AND R. WILSON, “Sequential Equilibria,” *Econometrica* 50 (1982), 863–94.

MAJUMDAR, D., AND A. SEN, “Ordinally Bayesian Incentive Compatible Voting Rules,” *Econometrica* 72 (2004), 523–40.

MIRALLES, A., “Cardinal Bayesian Allocation Mechanisms without Transfers,” *Journal of Economic Theory* 147 (2012), 179–206.

NEHRING, K., “The Veil of Public Ignorance,” *Journal of Economic Theory* 119 (2004), 247–70.

PALFREY, T. R., AND S. SRIVASTAVA, “Mechanism Design with Incomplete Information: A Solution to the Implementation Problem,” *The Journal of Political Economy* 97 (1989), 668–91.

RAE, D. W., “Decision-Rules and Individual Values in Constitutional Choice,” *The American Political Science Review* 63 (1969), 40–56.

ROCKAFELLAR, R. T., *Convex Analysis* (Princeton, NJ: Princeton University Press, 1970).

SAMUELS, S. M., “On the Number of Successes in Independent Trials,” *Annals of Mathematical Statistics* 36 (1965), 1272–8.

SCHMITZ, P. W., AND T. TRÖGER, “Garbled Elections,” CEPR Discussion Paper No. 195, 2006.

———, AND ———, “The (Sub-)Optimality of the Majority Rule,” *Games and Economic Behavior* 74 (2011), 651–65.

SHAPLEY, L. S., AND M. SHUBIK, “A Method for Evaluating the Distribution of Power in a Committee System,” *American Political Science Review* 48 (1954), 787–92.

VON-NEUMANN, J., AND O. MORGENSTERN, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).

WEYMARK, J. A., “Measurment Theory and the Foundations of Utilitarianism,” *Social Choice and Wefare* 25 (2005), 527–55.