Manipulation in Elections with Uncertain Preferences

Andrew McLennan*
School of Economics
Level 6, Colin Clark Building
University of Queensland
St Lucia, QLD 4072 Australia
mclennan@socsci.umn.edu

July 2010

Abstract

A decision scheme (Gibbard (1977)) maps profiles of strict preferences over a set of social alternatives to lotteries over the social alternatives. A decision scheme is weakly strategy-proof if it is never possible for a voter to increase expected utility (for some vNM utility function consistent with her ordinal preferences) by misrepresenting her preferences when her belief about the preferences of other voters is generated by a model in which the other voters are i.i.d. draws from a distribution over possible preferences. We show that if there are at least three alternatives, a decision scheme is necessarily a random dictatorship if it is weakly strategy-proof, never assigns positive probability to Pareto dominated alternatives, and is anonymous in the sense of being unaffected by permutations of the components of the profile.

Running Title: Manipulation in Elections

*This research was supported by Discovery Project DP0773324 of the Australian Research Council. I would like to thank an anonymous referee for useful remarks.
Journal of Economic Literature Classification Number D71.

**Keywords:** Gibbard-Satterthwaite theorem, strategy-proof, manipulation, voting, elections.
1 Introduction

Suppose we are given a finite set $V$ of social alternatives and $n$ voters. A preference profile is an $n$-tuple of strict individual preferences over $V$. A social choice function is a function whose domain is the set of preference profiles and whose range is $V$. The Gibbard-Satterthwaite theorem (Gibbard (1973), Satterthwaite (1975)) asserts that a social choice function must be dictatorial if there are at least three alternatives, the social alternative selected at a profile of preferences is never Pareto dominated for that profile, and the function is strategy-proof, which means that it is never possible for a voter to achieve a preferred outcome by reporting something other than her actual preference ordering.

When the electorate is large, voters typically have quite limited information about the preferences of other voters, so one should consider the possibility that a social choice function might be strategy-proof “in effect” if, in practice, voters are never able to manipulate because they lack sufficiently precise information. This paper develops a weakened notion of strategy-proofness that expresses this perspective. Our main results show that this weaker notion is still strong enough to imply a dictatorial conclusion.

Any mechanism combining the agents’ preferences in a nontrivial manner must depart from the framework of the Gibbard-Satterthwaite theorem in some respect, and for this reason the result is fundamental in the theory of mechanism design. The Gibbard-Satterthwaite theorem allows voters to have any strict preferences, but, for example, in the theory of matching (e.g., Roth and Sotomayor (1990)) agents are typically assumed to care only about whether they are matched and, if so, with whom. A domain restriction specifies a subset of the set of preference profiles. In the theory of voting the seminal concept of this sort is the notion of single peaked preferences, which leads to the median voter theorem (e.g., Black (1958)). Allowing the voters (but not the mechanism designer) to know each others’ preferences, and to behave with greater strategic sophistication, leads to the theory of Nash implementation pioneered by Maskin (1999).

Another relaxation of the Gibbard-Satterthwaite framework, that is more closely related to the work presented here, is to allow a random outcome. Indeed, tied elections are commonly resolved by coin flips, so this extension is very natural, and the Gibbard-Satterthwaite theorem would lose much of its force if there were electoral systems employing randomization that embodied democratic values. Gibbard (1977) defines a decision scheme to be
a function whose domain is the set of preference profiles and whose range is the set of probability distributions over $V$. Gibbard’s result (a precise explanation is given in Section 2) extends the Gibbard-Satterthwaite theorem to such schemes: if a decision scheme is strategy-proof in the sense that manipulation is never beneficial (when evaluated in terms of any von Neumann-Morgenstern utility consistent with the actual ordinal preference) and Pareto dominated alternatives never receive any probability, then the scheme must be a random dictatorship. Although random dictatorship is, in a certain sense, “fair” if each voter has the same probability of being the dictator, this result is generally considered to be quite negative because random dictatorship does not achieve the sorts of compromises one would hope to obtain from a democracy.

The main idea studied here might be thought of as a domain restriction, except that instead of imposing restrictions on the profiles that can occur, we impose restrictions on the voters’ beliefs about the profile. In effect, strategy proofness requires that a voter can never profitably manipulate when her belief assigns all probability to a particular profile. If truthful revelation is optimal at two beliefs, then it is optimal at any convex combination of those beliefs, so this amounts to the requirement that there is no belief whatsoever allowing profitable manipulation. There are various contexts in which this seems too strong because the agents are not fully informed about each others’ preferences, suggesting that nonmanipulability should only be required at beliefs about others’ preferences that might actually occur. Although the main results of this paper are negative, an important point is that there may be other mechanism design problems for which such a weakening of nonmanipulability allows interesting positive results. Indeed, a highly specific example is the framework of Bayesian implementation in which the given information, for both the agents and the mechanism designer, includes a prior distribution on the space of $n$-tuples of agent types.

We are particularly motivated by elections with many voters and more than two candidates such as the primaries used to select the parties’ candidates in the U.S. electoral system. In such contexts it is natural to suppose that voters will have quite limited information about each others’ preferences. Thus there arises the question of whether there might be an electoral system that is appealing in other respects and is immune to successful manipulation for all beliefs in a set that is large enough to encompass (perhaps in some approximate sense) any belief that does not embody more information than a voter might realistically be expected to possess.
This paper studies the effect of weakening nonmanipulability by requiring that profitable manipulation is impossible only for voters whose beliefs about the preferences of the other voters can be described by a model in which the other voters’ preferences are i.i.d. draws from a common distribution. Such a belief corresponds, roughly, to what a voter might believe after seeing a public opinion poll, and one would expect that any set of beliefs that was thought to be large enough to contain all “plausible” or “realistic” beliefs would include all beliefs of this form. Subject to certain mild qualifications described below, our main results show that the dictatorial conclusion is implied by this weak form of nonmanipulability and the other assumptions of Gibbard’s theorem. This seems to be a conceptually significant strengthening of Gibbard’s theorem, closing the door to what might otherwise have been a promising avenue of escape from its pessimistic conclusion.

Formally, a decision scheme is weakly strategy-proof if there is no voter, preference for that voter, von Neumann-Morgenstern utility consistent with that preference, and distribution over orderings of the alternatives, such that the voter can achieve a higher expected utility by manipulating when she regards the other voters’ preferences as i.i.d. draws from that distribution. The primary example of a weakly strategy proof decision scheme is a random dictatorship. In such a decision scheme there is a vector assigning a nonnegative weight to each agent, with the weights summing to one, and the probability of choosing a particular alternative is the total weight of the agents that have that alternative as their favorite. In effect, a voter is selected randomly, with the probabilities of being selected given by the weights, and that voter’s favorite alternative is the social choice. The anonymous random dictatorship is the random dictatorship in which all voters have the same weight $1/n$.

To illustrate the difference between strategy-proofness and weak strategy-proofness concretely, suppose that there are three voters, that from voter 1’s point of view the preferences of the other two voters are i.i.d. random variables, and that voter 1 can do better by manipulating when voter 2 has preference $P$ and voter 3 has preference ordering $Q$. The assumption that the preferences are i.i.d. implies that it is equally likely that voter 2 has preference $Q$ and voter 3 has preference $P$. In addition, if these two events are much more likely than both voters having preference $P$, then in turn both voters having preference $Q$ must be much more likely than either of these events. The fact that there is a profile at which voter 1 can profitably manipulate does not necessarily imply that there is a belief for voter 1 satisfying our assumption at which profitable manipulation is possible.
In fact when $n$ is large there are many decision schemes that never assign positive probability to Pareto dominated alternatives and are weakly strategy-proof, but not strategy-proof. For any particular voter with the sorts of beliefs we are allowing, and any particular profile, the voter regards all profiles obtained by permuting the preferences of the other voters as equally likely. For such a voter two decision schemes are effectively equivalent, if, for each profile, the two lotteries obtained by averaging the results of the two decision schemes over all permutations of the other voters’ preferences are the same. Since the number $(n - 1)!$ of permutations of the other voters grows rapidly with $n$, one can easily show, simply by counting equations and unknowns, that the set of decision schemes yielding the same averages as the anonymous random dictatorship has high dimension when $n$ is large.

These considerations suggest that we should focus attention on decision schemes that are anonymous in the sense that permuting the voter’s preferences does not affect the outcome. In most settings anonymity is regarded as desirable, or perhaps even one of the defining characteristics of democracy, and anonymity is certainly consonant with the spirit of our assumption concerning agents’ beliefs. For any decision scheme there is a derived anonymous decision scheme, which we will call its anonymization, in which the lottery assigned to a profile is the average of the lotteries assigned by the given decision scheme to the profiles obtained by permuting the components of the profile. The anonymization of an anonymous decision scheme is the scheme itself, so the process of anonymization partitions the decision schemes into equivalence classes, each of which has an anonymous central element that is the anonymization of every element of the class. A decision scheme never assigns probability to Pareto dominated alternatives if and only if its anonymization also has this property. If a decision scheme is weakly strategy-proof, then so is its anonymization.

We can now state our first main result: if a decision scheme is anonymous and weakly strategy-proof, and never assigns positive probability to Pareto dominated alternatives, and there are at least three alternatives, then it is the anonymous random dictatorship. This is proved in Section 3.

Since there is the additional hypothesis of anonymity, this is not, in a strict mathematical sense, a derivation of Gibbard’s conclusion from weaker assumptions. Thus there arises the question of what additional mechanisms are allowed if anonymity is dropped. There is an easy necessary condition: if there are at least three alternatives and a decision scheme is weakly strategy-proof and never assigns positive probability to Pareto dominated alternatives,
then its anonymization satisfies the assumptions of the theorem above and is consequently the anonymous random dictatorship. This condition is not sufficient: a decision scheme can be manipulable even though it’s anonymization is the anonymous random dictatorship because when an agent decides whether to manipulate, she knows her own component of the profile.

Do the additional possibilities resulting from dropping anonymity have enough value to undermine the negative conclusion of the result? One can imagine various arguments against anonymity: older voters might be accorded more weight because they are thought to be wiser; wealthier voters might be accorded more weight to promote stability, or less weight because they have greater access to other means of influencing the political process. By choosing the weights in a random dictatorship accordingly, one can seemingly accommodate such considerations. However, these examples of arguments for unequal weights mainly appeal to intuitions concerning the sorts of compromises the political system will eventually produce, so they provide little support for a view in which random dictatorships with diverse weights are superior to the anonymous random dictatorship.

Decision schemes whose anonymizations are the anonymous random dictatorship, but which are not themselves random dictatorships, seem to have few advantages over random dictatorships. From the point of view of an agent whose information is limited in the way we are imagining, such a mechanism differs from a random dictatorship only insofar as it uses the electorate itself as a source of randomness, and of course agents who have additional information may be able to manipulate.

In the discussion to this point we have assumed that the set of voters is fixed. More precisely, we have assumed that each voter has no uncertainty about who the other voters are, even if their preferences are uncertain. In actual elections voters often have quite imprecise knowledge concerning the pool of eligible voters, and in addition there is uncertainty about which ones will actually turn out. Thus it is very natural to consider models in which the size of the electorate is uncertain.

From the point of view of the sort of result described above, what qualities are desirable in a random model of the electorate? Since the conclusion is negative—an “impossibility theorem”—the result is more forceful if the conclusion is shown to hold even when voter’s beliefs are restricted to a relatively small set, since then the conclusion also holds when the voters’ beliefs are less restricted. A bit more subtly, the result is more forceful if the model of beliefs is a natural limit of other models, since the conclusion for
the other models can be derived from continuity. Finally, tractable models are preferred, of course. All these considerations strongly recommend the Poisson voting model developed by Meyerson (1998a,b, 2000, 2002) in which the number of voters with each preference is distributed according to a Poisson distribution, and these random variables are statistically independent. Section 4 extends the result described above to that setting.

2 Gibbard’s Theorem

Our analysis builds on several lemmas proved in Gibbard (1977), and the reader must refer to that source if she wishes to obtain a complete understanding of the proof. In order to create a package that is as seamless as possible we follow the notation and terminology of that paper quite closely. This section recapitulates the basic framework, and additional concepts from that paper are introduced in Section 3.

There is a nonempty finite set of alternatives $V$ whose elements are denoted by $x$, $y$, and $z$. A (strict) preference over $V$ is a complete transitive asymmetric binary relation on $V$. Such relations are denoted by $P$, $Q$, $P_k$, etc. A utility scale is a function $U : V \rightarrow \mathbb{R}$. The utility scale $U$ is said to fit the preference $P$ if more highly ranked alternatives give greater utility: for all $x, y \in V$, $U(x) > U(y)$ if and only if $xPy$. For any finite or countable set $A$, $\Delta(A)$ denotes the space of probability measures on $A$. A lottery is a probability measure on $V$. A utility is automatically interpreted (in the sense of von Neumann and Morgenstern) as extending linearly to $\Delta(V)$, so that $U(\mu) := \sum_{x \in V} U(x) \mu(x)$ whenever $\mu \in \Delta(V)$.

Society consists of $n$ voters, who are indexed by the integers $1, \ldots, n$. A profile is an $n$-tuple $\mathbf{P} = \langle P_1, \ldots, P_n \rangle$ assigning a preference to each voter. Let $\mathcal{P}$ be the set of profiles. For each $k = 1, \ldots, n$ let $\mathcal{P}_{-k}$ be the set of $(n - 1)$-tuples of preferences $\langle P_1, \ldots, P_{k-1}, P_{k+1}, \ldots, P_n \rangle$, thought of as configurations of preferences of the voters other than $k$. If $\mathbf{P} \in \mathcal{P}$ is given, $\mathbf{P}_{-k}$ will denote the $(n - 1)$-tuple obtained by dropping $P_k$. If $\mathbf{P}_{-k} \in \mathcal{P}_{-k}$ and $P_k'$ are given,

$$\langle \mathbf{P}_{-k}, P_k' \rangle = \langle P_1, \ldots, P_{k-1}, P_k', P_{k+1}, \ldots, P_n \rangle$$

is the profile obtained by combining these objects. If $\mathbf{P} \in \mathcal{P}$ and $P_k'$ are given,

$$\mathbf{P}_{/k} P_k' := \langle \mathbf{P}_{-k}, P_k' \rangle$$
is the profile obtained from $P$ by replacing $P_k$ with $P'_k$.

A decision scheme is a function

$$d : P \rightarrow \Delta(V).$$

We denote the probability assigned to alternative $x$ by the decision scheme at profile $P$ by $d(x, P)$, and for any $X \subseteq V$ we let $d(X, P) := \sum_{x \in X} d(x, P)$. We say that $d$ is a probability mixture of schemes $d_1, \ldots, d_m$ if there are positive numbers $\alpha_1, \ldots, \alpha_m$ with $\alpha_1 + \cdots + \alpha_m = 1$ such that

$$d(x, P) = \alpha_1 d_1(x, P) + \cdots + \alpha_m d_m(x, P)$$

for all alternatives $x$ and profiles $P$.

The decision scheme $d$ is potentially manipulable by $k$ at a profile $P$ if there is a utility scale $U$ that fits $P_k$ and a preference $P'_k$ such that $U(d(P/k, P'_k)) > U(d(P))$. We say that $d$ is manipulable if it is potentially manipulable by some voter at some profile, and otherwise it is strategy-proof. Note that a probability mixture of strategy-proof decision schemes is strategy-proof.

A lottery $\rho$ is Pareto optimal ex post for profile $P$ if $\rho(x) = 0$ for any alternative $x$ that is Pareto dominated insofar as there is another alternative $y$ such that $yP_i x$ for all $i$. The decision scheme $d$ is Pareto optimific ex post if, for each profile $P$, $d(P)$ is Pareto optimal ex post for $P$. If $d$ is a probability mixture of schemes $d_1, \ldots, d_m$, then $d$ is Pareto optimific ex post if and only if each $d_j$ is Pareto optimific ex post.

For a preference $P$, let $\varphi(P)$ be the top ranked alternative or favorite. A decision scheme $d$ is dictatorial, or a dictatorship, if there is a voter $k$ such that $d(\varphi(P_k), P) = 1$ for all $P \in \mathcal{P}$. A random dictatorship is a probability mixture of dictatorships.

Gibbard’s most general result asserts that a strategy proof decision scheme is a probability mixture of finitely many decision schemes, each of which is nonperverse (this concept is defined in the next section) and either duple, meaning that there are two alternatives that are the only alternatives receiving positive probability at any preference profile, or unilateral, meaning that it depends only on the preferences of a single voter. Gibbard credits Sonnenschein with the observation that, insofar as a duple decision scheme cannot be Pareto optimific ex post if there are three or more alternatives, and a unilateral decision scheme is Pareto optimific ex post if and only if it is dictatorial, it follows that:
Theorem 1 (Gibbard (1977)). If there are three or more alternatives and the decision scheme $d$ is strategy-proof and Pareto optimific ex post, then it is a random dictatorship.

3 The I.I.D. Model

A model of the electorate for voter $k$ is a probability measure $\beta \in \Delta(\mathcal{P}_- k)$. The decision scheme $d$ is potentially manipulable by $k$ at a model $\beta$ if there is a utility scale $U$ that fits $P_k$ and a preference $P'_k$ such that

$$U\left( \sum_{P_k \in \mathcal{P}_- k} d(P_k, P'_k) \beta(P_k) \right) > U\left( \sum_{P_k \in \mathcal{P}_- k} d(P_k, P_k) \beta(P_k) \right).$$

Let $\mathcal{O}$ be the set of all strict orderings of $V$. The model $\beta$ is identically and independently distributed (i.i.d.) if there is a $\sigma \in \Delta(\mathcal{O})$ such that $\beta(P) = \prod_{i \neq k} \sigma(P_i)$ for all $P_k \in \mathcal{P}_- k$. The decision scheme $d$ is strongly manipulable if there is a voter $k$ such that $d$ is potentially manipulable at some i.i.d. model of the electorate for $k$, and if this is not the case we say that $d$ is weakly strategy-proof.

Let $\mathbb{N} := \{0, 1, 2, \ldots \}$ be the nonnegative integers, and let $A := \mathbb{N}^\mathcal{O}$. An element of $A$ is called an anonymous profile because it specifies the number of voters with each preference ordering without attributing those preferences to specific individuals. For $a \in A$ let $|a| = \sum_{P \in \mathcal{O}} a_P$ be the total number of voters, and for $n = 0, 1, 2, \ldots$ let $A_n := \{ a \in A : |a| = n \}$. Let $\pi_n : \mathcal{P} \to A_n$ be the function defined by letting the component $\pi_{n,P}(P)$ be the number of $k$ such that $P_k = P$. The decision scheme $d$ is anonymous if $d(P)$ depends only on $\pi_n(P)$, so that there is a function $D_n : A_n \to \Delta(\mathcal{V})$ such that $d = D_n \circ \pi_n$. The anonymous random dictatorship is the decision scheme $d^*$ given by

$$d^*(x, P) := \frac{1}{n} \# \{ k : \varphi(P_k) = x \}.$$

Theorem 2. If there are three or more alternatives and the decision scheme $d$ is anonymous, weakly strategy-proof, and Pareto optimific ex post, then it is the anonymous random dictatorship.

The proof is developed in a sequence of lemmas. Fix a decision scheme $d$. A set $X \subset V$ heads a preference $P$ if $xPy$ for all $x \in X$ and $y \in V \setminus X$. If $d(X, P) = d(X, P / k P'_k)$ for all $k$, $P$, and $P'_k$ such that $X$ heads both $P_k$ and $P'_k$, then $d$ is said to be localized. The most innovative step in our argument is:
Lemma 1. If \( d \) is weakly strategy-proof and anonymous, then it is localized.

**Proof.** Fix a voter \( k \), a set \( X \subset V \), and \( P_k \) and \( P'_k \) such that \( X \) heads both \( P_k \) and \( P'_k \). We need to show that the probability of \( X \) is unaffected by whether \( k \) reports \( P_k \) or \( P'_k \). Let \( \pi_{-k} \) be \( \pi_{n-1} \) reinterpreted as a function with domain \( \mathcal{P}_{-k} \); for each \( P \), \( \pi_{-k,P}(\mathcal{P}_{-k}) \) is the number of \( j \) such that \( P_j = P \). Since we can interchange \( P_k \) and \( P'_k \), it suffices to show that \( M = \emptyset \) where \( M \) is the set of \( a \in A_{n-1} \) such that

\[
d(X, \langle \mathcal{P}_{-k}, P'_k \rangle) \neq d(X, \langle \mathcal{P}_{-k}, P_k \rangle)
\]

for some (hence all, because \( d \) is anonymous) \( \mathcal{P}_{-k} \) such that \( \pi_{-k}(\mathcal{P}_{-k}) = a \). Aiming at a contradiction, suppose that \( M \) is nonempty.

Below we will show how to find \( b \in M \) and \( \sigma \in \Delta(\mathcal{O}) \) such that the derived i.i.d. model of the electorate for \( k \), denoted by \( \beta \), has \( \beta(\mathcal{P}_{-k})/\beta(\mathcal{P}_{-k}) \) arbitrarily large for all \( \mathcal{P}_{-k} \) and \( \mathcal{P}'_{-k} \) such that \( \pi_{-k}(\mathcal{P}_{-k}) = b \) and \( \pi_{-k}(\mathcal{P}'_{-k}) = a \in M \setminus \{ b \} \). When these ratios are sufficiently large we have

\[
\sum_{\mathcal{P}_{-k} \in \mathcal{P}_{-k}} d(X, \langle \mathcal{P}_{-k}, P'_k \rangle) \beta(\mathcal{P}_{-k}) \neq \sum_{\mathcal{P}_{-k} \in \mathcal{P}_{-k}} d(X, \langle \mathcal{P}_{-k}, P_k \rangle) \beta(\mathcal{P}_{-k}).
\]

In this circumstance there is a utility scale \( U \) that fits \( P_k \), and which emphasizes the difference between \( X \) and \( V \setminus X \) while nearly disregarding differences between elements of \( X \) and between elements of \( V \setminus X \), to such an extent that either inequality (\(*\)) holds, or this inequality holds with \( P_k \) and \( P'_k \) interchanged, contradicting the assumption that \( d \) is weakly strategy-proof.

Let \( b \) be an extreme point of the convex hull of \( M \). Then \( b \in M \) because it is not in the convex hull of \( M \setminus \{ b \} \). Applying the separating hyperplane theorem to \( b \) and the convex hull of \( M \setminus \{ b \} \) gives a vector \( \ell \in \mathbb{R}^\mathcal{O} \) such that \( \langle \ell, b \rangle > \langle \ell, a \rangle \) for all \( a \in M \setminus \{ b \} \). For \( \alpha > 0 \) let \( \sigma_\alpha \in \Delta(\mathcal{O}) \) be the probability distribution in which the probability of each \( P \) is proportional to \( \alpha^{\ell_P} \), so that \( \sigma_\alpha(P) := \alpha^{\ell_P} / \sum_{P \in \mathcal{O}} \alpha^{\ell_P} \) and \( \sigma_\alpha(P)/\sigma_\alpha(P') = \alpha^{\ell_P}/\alpha^{\ell_{P'}} \) for all \( P, P' \in \mathcal{O} \). Let \( \beta_\alpha \) be the derived i.i.d. model of the electorate for \( k \); for each \( \mathcal{P}_{-k} \in \mathcal{P}_{-k} \), \( \beta_\alpha(\mathcal{P}_{-k}) = \prod_{i \neq k} \sigma_\alpha(P_i) \). Then for any \( \mathcal{P}_{-k}, \mathcal{P}'_{-k} \in \mathcal{P}_{-k} \) we have

\[
\frac{\beta_\alpha(\mathcal{P}_{-k})}{\beta_\alpha(\mathcal{P}'_{-k})} = \frac{\alpha^{\ell_{P_1}} \cdots \alpha^{\ell_{P_{k-1}}} \alpha^{\ell_{P_{k+1}}} \cdots \alpha^{\ell_{P_n}}}{\alpha^{\ell_{P'_1}} \cdots \alpha^{\ell_{P'_{k-1}}} \alpha^{\ell_{P'_{k+1}}} \cdots \alpha^{\ell_{P'_n}}} = \frac{\alpha^{\langle \ell, \pi_{-k}(\mathcal{P}_{-k}) \rangle}}{\alpha^{\langle \ell, \pi_{-k}(\mathcal{P}'_{-k}) \rangle}}.
\]

If \( \pi_{-k}(\mathcal{P}_{-k}) = b \) and \( \pi_{-k}(\mathcal{P}'_{-k}) = a \in M \setminus \{ b \} \), then \( \beta_\alpha(\mathcal{P}_{-k})/\beta_\alpha(\mathcal{P}'_{-k}) = \alpha^{\langle \ell, b \rangle}/\alpha^{\langle \ell, a \rangle} \to \infty \) as \( \alpha \to \infty \), as desired. \( \square \)
The remainder of the argument is a matter of marshalling tools developed in Gibbard (1977). We write \( xP!y \) to indicate that \( xPy \) and that \( x \) and \( y \) are adjacent in the ranking, so that for all \( z \notin \{x, y\} \), \( zPx \) if and only if \( zPy \). If this is the case, then \( P^y \) denotes the ranking obtained by interchanging \( x \) and \( y \) without changing the ranking of either in relation to any third alternative \( z \), and we say that \( P^y \) is obtained from \( P \) by switching \( x \) and \( y \). Given a profile \( P \) and a voter \( k \) with \( \varphi(P_k) \neq y \), let

\[
P^{ky} := \langle P_1, \ldots, P_{k-1}, P^y_k, P_{k+1}, \ldots, P_n \rangle.
\]

We say that \( d \) is pairwise responsive if \( d(z, P^{ky}) = d(z, P) \) for all distinct alternatives \( x, y, \) and \( z \), all voters \( k \), and all profiles \( P \) such that \( xP_k!y \). Of course if this is the case, then \( d(\{x, y\}, P^{ky}) = d(\{x, y\}, P) \) for all \( x, y, P, \) and \( k \) such that \( xP_k!y \).

**Lemma 2.** \( d \) is localized if and only if it is pairwise responsive.

**Proof.** This follows from Lemma 1 (p. 672) of Gibbard (1977). \( \square \)

Given a profile \( P \) and a voter \( k \) with \( \varphi(P_k) \neq y \), the effect under \( d \) of \( k \)'s switching \( y \) upward is

\[
\varepsilon^y_k(d, P) := d(y, P^{ky}) - d(y, P).
\]

The decision scheme \( d \) is nonperverse if \( \varepsilon^y_k(d, P) \geq 0 \) for every \( P, k, \) and \( y \neq \varphi(P_k) \). If \( P \) is an ordering and \( x, y \in V, P^\uparrow\{x, y\} \) is the ordering of \( \{x, y\} \) obtained by restricting \( P \) to this set. For \( P \in \mathcal{P} \) the derived profile of preferences over \( \{x, y\} \) is

\[
P^\uparrow\{x, y\} := \langle P^\uparrow\{x, y\}, \ldots, P_n^\uparrow\{x, y\} \rangle.
\]

We say that \( d \) is pairwise isolated if

\[
\varepsilon^y_k(d, P) = \varepsilon^y_k(d, P')
\]

for all \( P, P', x, \) and \( y \) such that \( P^\uparrow\{x, y\} = P'^\uparrow\{x, y\} \) and all \( k \) such that \( P_k = P'_k \) and \( xP_k!y \). The decision scheme \( d \) is decomposable if, for any fixed \( k, x, \) and \( y \) with \( x \neq y \), there are functions \( \gamma \) and \( \delta \) such that for all \( P \) with \( xP_k!y \),

\[
\varepsilon^y_k(d, P) = \gamma(P^\uparrow\{x, y\}) + \delta(P_k).
\]
Lemma 3. If \( d \) is localized, then it is pairwise isolated and decomposable.

Proof. This is Lemma 3 (p. 673) of Gibbard (1977).

For a profile \( \mathbf{P}, \varphi(\mathbf{P}) := \langle \varphi(P_1), \ldots, \varphi(P_n) \rangle \). We say that \( d \) depends only on favorites if \( d(\mathbf{P}') = d(\mathbf{P}) \) for all profiles \( \mathbf{P} \) and \( \mathbf{P}' \) such that \( \varphi(\mathbf{P}') = \varphi(\mathbf{P}) \).

Lemma 4. If \( d \) is localized and Pareto optimific ex post, and there are three or more alternatives, then \( d \) depends only on favorites. If, in addition, \( d \) is anonymous, then it is the anonymous random dictatorship.

Proof. First consider particular \( k, x, \) and \( y \) with \( x \neq y \). Since \( d \) is localized, it is decomposable; let \( \gamma \) and \( \delta \) be the functions given by the definition of decomposability. Since there are three distinct alternatives, for some \( z \notin \{x, y\} \) there are profiles \( \mathbf{P} \) with \( \varphi(P_i) = z \) for all \( i \). Since \( d \) is Pareto optimific ex post, \( d(z, \mathbf{P}) = 1 \) for any such \( \mathbf{P} \). By allowing \( \mathbf{P} \) to vary in the set of such profiles we can deduce that \( \gamma \) is identically zero, and that \( \delta(P_k) = 0 \) whenever \( xP_k!y \) and \( x \neq \varphi(P_k) \). Since \( d \) is pairwise responsive, and it is possible to move between any two \( \mathbf{P} \) and \( \mathbf{P}' \) with \( \varphi(\mathbf{P}) = \varphi(\mathbf{P}') \) through a sequence of switches that do not affect the vector of favorites, \( d \) depends only on favorites.

It now follows that there are numbers \( \epsilon_k(x, y) \in [0, 1] \) such that \( \epsilon_k(x, y) \) whenever \( x = \varphi(P_k) \) and \( xP_k!y \). Consider a profile \( \mathbf{P} \) in which all voters rank \( x \) first and \( y \) second. Since \( d \) is Pareto optimific ex post, by switching \( x \) with \( y \) one voter at a time we obtain

\[
\epsilon_1(x, y) + \cdots + \epsilon_n(x, y) = 1.
\]

If \( d \) is anonymous, then \( \epsilon_k(x, y) \) does not depend on \( k \), so \( \epsilon_k(x, y) = 1/n \) for all \( k, x, \) and \( y \).

Theorem 2 follows from Lemmas 1 and 4.

4 The Poisson Model

In earlier sections a decision scheme was a function whose argument was an assignment of preferences to a fixed set \( \{1, \ldots, n\} \) of “names.” When the set of voters is variable, any attempt to keep track of names would be cumbersome at best, and irrelevant to our aims, so we adopt a definition
that embeds the assumption of anonymity. An extended anonymous decision scheme (EADS) is a function

\[ D : A \to \Delta(V) \]

from anonymous profiles to lotteries. Fix such a \( D \). Let \( D(x, a) \) be the probability that \( x \) is chosen when the anonymous profile is \( a \), and for \( X \subset V \) let \( D(X, a) := \sum_{x \in X} D(x, a) \).

As in the last section, we wish to develop a mild notion of strategy proofness based on restrictions on the beliefs a voter may hold about the preferences of other voters. For a random variable taking values in \( \mathbb{N} \) that is distributed according to the Poisson distribution with mean \( \mu \), the probability that the variable takes on value \( m \) is

\[ f_\mu(m) := \mu^m e^{-\mu}/m!. \]

For \( \lambda \in (0, \infty)^\mathcal{O} \) and \( a \in A \) let

\[ F_\lambda(a) := \prod_{P \in \mathcal{O}} f_\lambda_P(a_P). \]

Then \( F_\lambda \) specifies a model of the electorate in which the numbers \( a_P \) of voters with each preference are independent random variables and each \( a_P \) has a Poisson distribution with mean \( \lambda_P \).

Poisson models of elections and more general games have been studied extensively by Meyerson (1998a,b, 2000, 2002). The Poisson distribution with mean \( \mu \) is the limit as \( N \to \infty \) of the distribution of the number of heads among \( N \) independent coin flips, each of which comes up heads with probability \( \mu/N \). Insofar as Poisson models are limits of related or more general models, results such as Theorem 3 below imply, by continuity, corresponding results for “nearby” models. In part because they are limits, Poisson models tend to be especially tractable.

A particularly pleasant property, which we take as the basis of our analysis, is called environmental equivalence. Suppose that a voter believes that the probability of being one of \( m+1 \) voters with preference \( P \) is proportional to \( m+1 \) times the probability that there are \( m+1 \) voters with preference \( P \). Then the probability, conditional on being a voter with preference \( P \), that there are \( m \) other voters with preference \( P \), should be

\[
\frac{(m+1)f_\lambda_P(m+1)}{\sum_{j=1}^\infty j f_\lambda_P(j)} = \frac{\lambda_P^{m+1}/m!}{\sum_{j=1}^\infty \lambda_P^j/(j-1)!} = \frac{\lambda_P^m/m!}{e^{\lambda_P}} = f_\lambda_P(m).
\]
Thus the voter’s belief concerning the number of other voters with preference \( P \) coincides with the given distribution of the total number of voters with this preference. Since the numbers \( a_P \) are statistically independent, being a voter with preference \( P \) conveys no information about the number of voters with any other preference. Therefore the voter’s belief about the rest of the electorate should coincide with the model’s description of the entire electorate. These calculations are heuristic, appealing to intuitions concerning the perspective of one member of an infinite pool of potential voters, each of whom is chosen with infinitesimal probability, but they can be made precise by taking the limit of a sequence of models in which, for each \( P, N_P \) potential voters with preference \( P \) each have probability \( \lambda_P/N_P \) of being included in the electorate.

An extended model of the electorate is a probability measure \( B \in \Delta(\mathcal{A}) \). For each \( P \) let \( e_P \) be the element of \( \mathcal{A} \) whose \( P \)-coordinate is 1 and whose other coordinates are 0. We say that \( D \) is potentially manipulable by preference ordering \( P \) at \( B \) if there is a utility scale \( U \) that fits \( P \) and a preference \( P' \) such that

\[
U\left( \sum_{a \in \mathcal{A}} D(a + e_{P'})B(a) \right) > U\left( \sum_{a \in \mathcal{A}} D(a + e_P)B(a) \right)
\]

\((**).\)

The EADS \( D \) is strongly manipulable if there is a preference \( P \) such that \( D \) is potentially manipulable by \( P \) at some Poisson model \( F_\lambda \), and if this is not the case we say that \( D \) is weakly strategy-proof.

We say that \( D \) is Pareto optimific ex post if \( D(x, a) = 0 \) whenever there is \( y \) such that \( y P x \) for all \( P \) with \( a_P > 0 \). The extended anonymous random dictatorship is the EADS \( D^* \) given by

\[
D^*(x, a) := \frac{1}{|\mathcal{A}|} \sum_{\varphi(P) = x} a_P.
\]

**Theorem 3.** If there are three or more alternatives and the EADS \( D \) is weakly strategy-proof and Pareto optimific ex post, then it is the extended anonymous random dictatorship.

We say that the EADS \( D \) is localized if \( D(X, a + e_P) = D(X, a + e_{P'}) \) for all \( a \in \mathcal{A} \) and \( P, P' \in \mathcal{O} \) such that \( X \) heads both \( P \) and \( P' \). For each \( n = 1, 2, \ldots \) there is a unique anonymous decision scheme \( d_n \) such that \( d_n = D \circ \pi_n \). Clearly \( D \) is localized if and only if each \( d_n \) is localized, \( D \) is
Pareto optimific ex post if and only if each \( d_n \) is Pareto optimific ex post, and \( D \) is the extended random anonymous dictatorship if and only if each \( d_n \) is the anonymous random dictatorship for that \( n \). Therefore Theorem 3 follows from Lemma 4 and the following analogue of Lemma 1.

**Lemma 5.** If \( D \) is weakly strategy proof, then it is localized.

**Proof.** Fixing a nonempty \( X \subset V \) and \( P \) and \( P' \) such that \( X \) heads both \( P \) and \( P' \), let \( M \) be the set of \( a \in A \) such that

\[
D(X, a + e_{P'}) \neq D(X, a + e_P).
\]

Our goal is to show that \( M = \emptyset \). Supposing otherwise, let \( n \) be the minimum value of \( |a| \) for \( a \in M \), let \( L \) be the set of \( a \in M \) such that \( |a| = n \), and let \( N := L \cup \{ a \in A : |a| > n \} \). Then \( L \subset M \subset N \).

Let \( B \) be the convex hull of \( L \), and let \( C \) be the convex hull of \( N \). Then \( B \) is the convex hull of its extreme points, each of which is an element of \( L \). Let \( b \) be one of these extreme points. Then \( b \) is also an extreme point of \( C \) because any representation of \( b \) as convex combination of elements of \( C \) must assign positive weight only to points whose components sum to \( n \). The separating hyperplane theorem gives a vector \( \ell \in \mathbb{R}^O \) such that \( \langle \ell, b \rangle > \langle \ell, a \rangle \) for all \( a \in N \setminus \{b\} \). Since \( a \in N \) whenever \( a_Q \geq b_Q \) for all \( Q \), all components of \( \ell \) are negative.

For some \( \alpha > 0 \) let \( \lambda \) be the vector with components \( \lambda_Q := \alpha \ell_Q \). For \( a \in A \) we have \( \prod_Q \alpha_a^{a_Q} = \alpha^{\sum_a \ell_a a_Q} = \alpha^{\langle \ell, a \rangle} \) and

\[
F_\lambda(a) = \prod_Q \lambda_Q^{a_Q} e^{-\lambda_Q/a_Q!} = \alpha^{\langle \ell, a \rangle} e^{-\sum_Q \lambda_Q} \prod_Q a_Q!.
\]

In particular

\[
\frac{F_\lambda(a)}{F_\lambda(b)} = \left( \prod_Q b_Q!/a_Q! \right) \frac{\alpha^{\langle \ell, a \rangle}}{\alpha^{\langle \ell, b \rangle}}.
\]

For each \( a \in N \setminus \{b\} \) this quantity goes to zero as \( \alpha \to \infty \), but in fact a stronger statement is true and relevant. Since the components of \( \ell \) are negative, \( \langle \ell, a \rangle \) is bounded above by a negative multiple of \( |a| \). The number of \( a \in A \) with \( |a| = n' \) is a polynomial function of \( n' \), and \( \prod_Q b_Q!/a_Q! \) is bounded above by \( \prod_Q b_Q! \). Therefore

\[
\sum_{a \in M \setminus \{b\}} \frac{F_\lambda(a)}{F_\lambda(b)} \leq \sum_{a \in N \setminus \{b\}} \frac{F_\lambda(a)}{F_\lambda(b)} \to 0 \quad \text{as} \quad \alpha \to \infty,
\]

16
\[
\sum_{a \in A} D(X, a + e_{P'}) F_\lambda(a) \neq \sum_{a \in A} D(X, a + e_P) F_\lambda(a)
\]

when \( \alpha \) is sufficiently large, in which case there is a utility scale \( U \) that fits \( P \) such that either (**) or (**) with \( P \) and \( P' \) interchanged holds when \( B = F_\lambda \). This contradiction of the assumption that \( D \) is weakly strategy-proof completes the proof. \( \square \)

References


