# Generic Difference of Expected Vote Share and Probability of Victory Maximization in Simple Plurality Elections with Probabilistic Voters

John W. Patty[1]
Department of Social and Decision Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
*jwpatty@cmu.edu*

July 29, 2004

[1]Discussions with Kim Border and Bill Zame played a pivotal role in the completion of this work. In addition, the author gratefully acknowledges the helpful comments of Jeff Banks, John Duggan, Mark Fey, Richard McKelvey, Tom Palfrey and especially the thoughtful comments and suggestions of an anonymous referee. Financial support for this work was provided by the Alfred P. Sloan Foundation as well as the California Institute of Technology.

**Abstract**

In this paper I examine single member, simple plurality elections with $n \geq 3$ probabilistic voters and show that the maximization of expected vote share and maximization of probability of victory are "generically different" in a specific sense. More specifically, I first describe *finite shyness* (Anderson and Zame (2000)), a notion of genericity for infinite dimensional spaces. Using this notion, I show that, for any policy $x^*$ in the interior of the policy space and any candidate $j$, the set of $n$-dimensional profiles of twice continuously differentiable probabilistic voting functions for which $x^*$ simultaneously satisfies the first and second order conditions for maximization of $j$'s probability of victory and $j$'sexpected vote share at $x^*$ is finitely shy with respect to the set of $n$-dimensional profiles of twice continuously differentiable probabilistic voting functions for which $x^*$ satisfies the first and second order conditions for maximization of $j$'s expected vote share.

# 1 Introduction

In this paper, I examine the question of equivalence of two different objective (or payoff) functions that political candidates may seek to maximize in an election: expected vote share or the probability of victory. I restrict attention to single winner, simple plurality elections with probabilistic voters and inquire as to whether optimal candidate strategies and equilibrium policy positions are different under these two objective functions. The main finding of this paper is that expected vote share and probability of victory are "generically" different in the sense that satisfaction of the first and second order conditions for maximization of expected vote share by an electoral platform generally does not imply satisfaction of the first and second order conditions for maximization of probability of victory.

The question of equivalence between different candidate objectives, first seriously studied in the 1970s (Aranson, Hinich, and Ordeshook, (1974), Hinich (1977), and Ledyard (1984)), has been the subject of renewed interest recently (Duggan (2000) and Patty (2000), (2001)). At issue is whether candidates who seek to maximize their vote share should adopt the same strategies as candidates who seek to maximize the probability of winning the election. In this paper I prove that the answer to this question for single member, simple plurality elections with probabilistic voters is, in a precise sense, "almost always" no.

There are two types of equivalence that have interested scholars of electoral strategy, *best response* and *equilibrium* equivalence. If the optimal strategies of the candidates are identical under the two objective functions, regardless of their opponents' policy choices, then the objective functions are said to exhibit best response equivalence. Equilibrium equivalence of two objectives holds if the two objectives yield identical sets of Nash equilibria. This paper speaks to both types of equivalence. More to the point, the paper illustrates that either type of equivalence between vote-maximization and probability of victory maximization is nongeneric. In other words, one can confidently expect candidate behavior to differ under vote-maximization and probability of victory maximization, regardless of whether the object of interest is individual incentives or equilibrium behavior.

The main point of this paper's results is that the optimal strategies for expected-vote-maximizing and probability-of-victory-maximizing candidates usually differ. This result is of theoretical and substantive importance for a number of reasons: first, there is no reason to assume *a priori* that the predictions of models of electoral competition are invariant to which of these two objectives motivate candidates' choices of platforms. Secondly, a probability of victory-maximizing candidate will not generally choose a platform in a manner such that the expected behaviors of all voters are treated "equally": the responsiveness of a voter's behavior is weighted by the probability of his or her vote being pivotal in the election when the candidate calculates the marginal benefit of a deviation in platforms. Finally, a pre-election poll of expected vote choices is a sufficient statistic for expected vote share (so long as voters respond to the poll truthfully) – these results indicate that there is no reason to assume without further restrictions that such a poll also provides a sufficient statistic for the candidates' probabilities of winning the election.

A review of the relevant literature is provided in Section 2. The model is defined in Section

3. In Section 4 we present a notion of genericity for infinite dimensional spaces, *shyness*, due to Hunt, Sauer, and Yorke (1992), and recently generalized by Anderson and Zame (2000). In Section 5 I present several lemmas and the main result of the paper: generically, a policy that satisfies the first and second order necessary conditions for maximization expected vote share does not satisfy the first and second order necessary conditions for maximization of probability of victory. The final section concludes.

# 2 Related Work

Aranson, Hinich, and Ordeshook (1974) offer an equivalence result which rests on assumptions regarding perturbations of the candidate's objective functions, perhaps representing forecast errors. Their result, however, requires that these forecast errors are unbiased and, more importantly, that the errors are uncorrelated with the strategies chosen by the candidates. As the authors point out, this assumption is untenable, since the value of the objective functions (even after the errors are taken into account) must fall between zero and one. A second equivalence result obtained by Aranson, *et al.* requires that the votes received in a two candidate election be distributed according to a multivariate normal distribution. This obviously requires that negative vote totals be a positive probability event. Aranson, *et al.* were unable to offer any equivalence results between expected plurality and probability of victory based on assumptions regarding the primitives of the model.

Hinich (1977) provides justification for examining expected vote share in place of probability of victory which depends only on the Central Limit Theorem. Hinich's equivalence result states that the two objective functions converged in 2 candidate elections without abstention. This finding was extended by Ledyard (1984) to include 2 candidate elections in which abstention is allowed.

Patty (2001) examines expected vote share maximization, expected plurality maximization, and maximization of probability of victory and provides counterexamples to Hinich's and Ledyard's results as well as providing sufficient conditions for best response equivalence in two candidate elections without abstention. Duggan (2000) examines the question of local equilibrium equivalence in two candidate elections without abstention. Restricting attention to a voter behavior rationalizable by an additive utility bias model of random utility maximization, Duggan proves that a strengthened version of local concavity of voter preferences at a policy profile is a sufficient condition for local equilibrium equivalence between maximization of expected vote share and maximization of probability of victory. Patty (2000) provides a related notion of local equilibrium equivalence and essentially extends Duggan's findings to general models of probabilistic voting as well as elections with more than two candidates.

To date, research on the question of equivalence has successfully provided several sufficient conditions for both best response and local equilibrium equivalence. The literature has been relatively silent, however, on the question of necessary conditions. Indeed, it is the author's impression that most scholars consider the occurrence of best response equivalence to be a rare event. This intuition has not yet been formalized in the literature. This paper attempts to offer a rigorous examination of this issue within models of probabilistic voting and single member, simple plurality elections.

# 3 The Model

Let $N$ denote a finite set of voters, with $|N| = n \geq 3$, and $J$ denote the set of candidates, with the cardinality of $J$ being denoted as usual by $|J|$. Each candidate $j \in J$ simultaneously chooses a point $x_j$ in some compact policy space $X \subset \mathbb{R}^K$, with $K < \infty$, possessing nonempty interior. I denote a $J$-dimensional vector of policy proposals by $x$ and the space of all such vectors of policy proposals by $Y = X^{|J|}$. The vector of all announced policies, other than the policy announced by candidate $j$, is denoted by $x_{-j}$, and the space of all such vectors by $Y_{-j}$.

Each voter $i$ chooses one candidate, denoted by $a_i \in J$.[1] The vector of all choices, $(a_1, \ldots, a_N)$, is denoted by $a$. The space of all such vectors is denoted by $A$. Each candidate $j$ possesses an *objective function* $u_j : A \to \mathbb{R}$. For any $a \in A$ and $j \in J$, I denote the *vote total* of candidate $j$ by $v_j(a) = \sum_{i=1}^{N} \mathbf{1}[a_i = j]$ and let $w(a) \in \{j \in J | v_j(a) \geq \max_{l \in J} v_l(a)\}$ denote the winning candidate at $s$. In the case of a tie, the winner is assumed to be determined by a fair lottery between all candidates $j$ for which $v_j(a) = \max_{l \in J} v_l(a)$. I denote the set of such candidates by $W(a)$. Thus, I am restricting attention to single winner, simple plurality rule systems with a fair tie-breaking rule.

This paper considers elections with probabilistic voters (see Coughlin (1992) for an explication and survey of the theory of probabilistic voting). Accordingly, each voter $i \in N$ is characterized by a twice continuously differentiable *response function*, $p_i : Y \to \Delta(J)$, where $\Delta(J)$ denotes the $|J| - 1$ dimensional simplex: the set of $|J|$-dimensional vectors $\pi$ for which $\sum_{j \in J} \pi^j = 1$ and $\pi^j \geq 0$ for all $j \in J$. I denote the probability an alternative $j \in J$ receives voter $i$'s vote, conditional on policy proposal vector $x$, by $p_i^j(x)$. I denote the vector of all voters' response functions by $p$.

I assume that each $p_i(x)$ characterizes an independent multinomial random variable $a_i(x)$, meaning that, given a policy profile $x \in Y$, all voters' votes are independent. This is stated formally below.

**Assumption 1 (Independence)** *Conditional on a vector of policy proposals, $x \in Y$, the set of $a_i(x)$ are independent random variables, each distributed according to $p_i(x)$, respectively, for all $i \in N$.*

I now use the set of response functions, $p$, to define two candidate objective functions, expected vote share and probability of victory. Given any profile of policy proposals $x \in Y$, any vector of response functions $p$, and for any vector of vote choices $a$, we write $\Pr[a|p(x)] = \prod_{i \in N} p_i^{a_i}(x)$ to denote the probability that the vote vector $a$ is realized.

Given opponents' pure strategies $x_{-j}$, an *expected vote share maximizing candidate* $j \in J$

[1] I do not examine abstention in this paper.

seeks to maximize

$$V_j(x) = \frac{1}{N} \sum_{i=1}^{N} p_i^j(x)$$

and a *probability of victory maximizing candidate* $j \in J$ seeks to maximize

$$R_j(x) = \sum_{a \in A} \left( \frac{1}{|W(a)|} \mathbf{1}[j \in W(a)] \Pr[a|p(x)] \right).$$

I define an electoral game as $\Gamma = (J, N, X, p, u)$, where $u$ is a $J$-dimensional vector of candidate objective functions such that $u_j \in \{V_j, R_j\}$ for each candidate $j \in J$.[2]

In words, best response equivalence holds whenever two objective functions prescribe an identical optimal (pure) strategy regardless of the strategieschosen by the opponents.[3] Such equivalence is essentially a decision-theoretic concern, as the strategic effects of other players' motivations are inconsequential to the player in question. A second, weaker, form of equivalence is equilibrium equivalence. Equilibrium equivalence holds whenever the set of Nash equilibria under two different objective functions are identical. It is straight-forward to show that best response equivalence implies equilibrium equivalence, so that equilibrium equivalence is a *necessary* condition for best response equivalence.[4] This paper offers insight into both of these questions in the case where each voter's behavior is a twice continuously differentiable function of the policy choices of the candidates by examining the satisfaction of the necessary first and second order conditions for maximization of the two objectives.

# 4 Shyness and Finite Shyness

Finite shyness, as defined by Anderson and Zame (2000), provides a rigorous notion of genericity in infinite-dimensional spaces.[5] It is intended to behave in ways similar to measure-theoretic notions of genericity (*i.e.*, a notion of "almost everywhere") in finite dimensional spaces. The space of interest in this paper is the space of twice continuously differentiable functions from a compact set $Y$ to the $n$-fold Cartesian product of $|J| - 1$ dimensional simplices, $\Delta(J)^n$. This space is infinite-dimensional, leading to our interest in the notion of finite shyness. I now proceed to define this notion.

For any finite dimensional subspace $V \subset X$, let $\lambda_V$ denote Lebesgue measure on $V$ and, analogously, write $\lambda_{\mathbf{R}^k}$ for Lebesgue measure on $\mathbf{R}^k$.[6]

[2]The candidates are not required to share the same objective: some may maximize expected vote while others maximize probability of victory.

[3]Throughout this paper, attention is restricted to pure strategies by the candidates. A discussion of best response equivalence in the space of mixed strategies is contained in Chapter 2 of Patty (2000).

[4]For a more detailed discussion of this, see Aranson, Hinich, and Ordeshook (1974), p. 144-145.

[5]Finite shyness is an extensions of the notion of shyness, as defined by Hunt, Sauer, and Yorke (1992). Finite shyness is a stronger version of shyness.

[6]As noted by Anderson and Zame (2000) (p.13, footnote 11), for any finite dimensional space $V$ there exists a continuous linear isomorphism $T : V \to \mathbb{R}^k$ for some positive integer $k$. Given $T$, one can define $\lambda_V(A) =$

**Definition 1** *Let $Q$ be a topological vector space and let $U$ be a convex subset of $Q$ that is completely metrizable in the relative topology induced by $Q$. A Borel subset $E \subset U$ is* finitely shy in (*or* relative to) $U$ *if there is a finite-dimensional subspace $V \subset Q$ such that $\lambda_V(U + a) > 0$ for some $a \in Q$ and $\lambda_V(E + q) = 0$ for every $q \in Q$. An arbitrary subset $F \subset Q$ is finitely shy in $U$ if it is contained in a finitely shy Borel set. If $E$ is finitely shy in $U$, then $U \setminus E$ is referred to as* finitely prevalent.

A useful fact is that the finite union of finitely shy sets is itself finitely shy.

Before presenting the analysis and results I note that, throughout the paper, the ambient topological vector space (i.e., the topological vector space $Q$ in the above definitions) is taken to be the space of twice continuously differentiable functions from $Y$ to $(\mathbb{R}^{|J|})^n$, endowed with the topology of $C^2$ uniform convergence.[7] This space, which is complete, separable, and metrizable (Mas-Colell, (1985), p.50), is denoted by $\mathcal{C}^2$ throughout the paper. The space of $n$-dimensional vectors of twice continuously differentiable response functions is denoted by $P(Y)$, a closed subset of $\mathcal{C}^2$.

# 5 Analysis and Results

In this section it is first shown that, for any policy profile $x^*$ in the interior of $Y$ and any candidate $j$, the set of $n$-dimensional vectors of twice differentiable response functions that lead to simultaneous satisfaction of the first and second order necessary conditions for maximization of $V_j$ and $R_j$ at $x^*$ is shy in the set of $n$-dimensional vectors of twice differentiable response functions that satisfy the first and second order necessary conditions for maximization of $V_j$ at $x^*$. This then immediately implies (the much weaker result) that the set of $n$-dimensional vectors of twice differentiable response functions that exhibit best response equivalence is shy in the set of all $n$-dimensional vectors of twice differentiable response functions.[8]

The results are stated in what may appear to be a strange fashion. In particular, a profile of platforms is fixed and the sets of response functions which exhibit equivalence *at that point* are examined. This method is motivated by application; typically, the question of equivalence is dealt with when a modeler seeks to verify that, for example, the equilibrium derived under one objective function is also an equilibrium under the other objective. Thus, the results provided here state that, supposing that $x^* \in Y$ satisfies the necessary conditions to be a best response

$\lambda_{\mathbf{R}^k}(T(A))$ for each Borel set $A \subset V$. While this derived measure depends on the choice of isomorphism $T$, all measures derived in this way are mutually absolutely continuous, so that for two isomorphisms $T$ and $T'$, $\lambda_{\mathbf{R}^k}(T(A)) = 0 \Rightarrow \lambda_{\mathbf{R}^k}(T'(A)) = 0$ for any Borel set $A \subset V$. We are concerned only with sets of Lebesgue measure zero, so any choice of isomorphism $T$ is without loss of generality for the purposes of this paper.

[7] Denoting the $i^{th}$ derivative of a function $f$ by $f^i$, the topology of $C^r$ uniform convergence is the topology generated by the semimetric

$$d^r(f,g) = \max_{0 \le i \le r} \left[ \sup_{y \in Y} ||f^i(y) - g^i(y)|| \right],$$

where $||x|| = (\sum_s x_s^2)^{1/2}$ denotes the usual Euclidean metric.

[8] I thank a referee for clarifying my thinking regarding, and the exposition of, this point.

under $V$, it is "generally not the case" that $x^*$ also satisfies the necessary conditions to be a best response under $R$.

## 5.1 Generic Failure of Equivalence

For any electoral game with differentiable response functions $p$, any candidate $j \in J$, and any policy profile $x \in Y$, the first derivative of candidate $j$'s expected vote with respect to $j$'s policy choice is

$$D_{x_j} V_j(x) = \sum_{i \in N} D_{x_j} p_i^j(x).$$

Define the pivot probability of voter $i$ with respect to candidate $l$, given a policy profile $x \in Y$ and other voters' response functions $p_{-i}$, as

$$\delta_i^l(p_{-i}(x)) = \sum_{a \in \mathcal{D}(i;l)} \left[ \frac{1}{|W(a)|} \prod_{j \neq i} p_j^{a_j}(x) \right], \tag{1}$$

where $\mathcal{D}(i;j) \subset A$ denotes the set of vote vectors in which voter $i$ is decisive (or pivotal) for candidate $j$. That is, $\mathcal{D}(i;j)$ is the set of outcomes in which voter $i$'s vote for candidate $j$ either created a tie between $j$ and some other candidate(s) or broke a tie between $j$ and some other candidate(s). The following result (proved in the appendix) uses the pivot probability to express the first derivative of a candidate's probability of victory with respect to her own policy choice.

**Lemma 1** *For any electoral game with differentiable response functions $p$, any candidate $j \in J$, and any policy profile $x \in Y$,*

$$D_{x_j} R_j(x) = \sum_{i \in N} \delta_i^j(p_{-i}(x)) D_{x_j} p_i^j(x).$$

For any point $x^* \in Int(Y)$, define $P_V(x^*) \subset P(Y)$ as the set of $n$-dimensional vectors of twice continuously differentiable response functions such that, for all $j \in J$,

$$DV_j(x^*) = \sum_{i=1}^{n} Dp_i^j(x^*) = 0$$

and

$$D^2 V_j(x^*) = \sum_{i=1}^{n} D^2 p_i^j(x^*) \text{ is negative semidefinite (n.s.d),}$$

where $Dp_i^j(x^*)$ denotes the evaluation at $x^*$ of the first derivative of voter $i$'s probability of voting for candidate $j$ with respect to candidate $j$'s policy announcement and where $D^2 p_i^j(x^*)$ denotes the evaluation at $x^*$ of the matrix of second partial derivatives of voter $i$'s probability of voting for candidate $j$ with respect to candidate $j$'s policy announcement. Similarly, let

$P_R(x^*)$ denote the set of $n$-dimensional vectors of twice continuously differentiable response functions such that, for all $j \in J$,

$$DR_j(x^*) = \sum_{i=1}^{N} \delta_i^j(p_{-i}(x^*))Dp_i^j(x^*) = 0.$$

and

$$D^2R_j(x^*) \text{ is negative semidefinite.}$$

Finally, let $P_{V,R}(x^*)$ denote the intersection of $P_V(x^*)$ and $P_R(x^*)$.

Before continuing, it should be noted that, while the definition of the set takes $x^*$ as an argument, this is appropriate for the purposes of this paper in two respects: first, the main result of the paper is that any pure strategy that satisfies the first and second order conditions for maximization of expected vote share maximization is extremely unlikely to also satisfy the first and second order conditions for maximization of probability of victory and, second, the results do not use any special characteristics of $x^*$ other than the fact that it is in the interior of $Y$.[9]

The main result in this section is that $P_{V,R}(x^*)$ is finitely shy in $P_V(x^*)$ for any $x^* \in Int(Y)$. First, several lemmas are proved. The first two lemmas jointly demonstrate that the set of function profiles in $P_{V,R}(x^*)$ such that there exists at least one voter $i$ and one candidate $j$ for which

$$\delta_i^j(p_{-i}(x^*))Dp_i^j(x^*) \neq 0$$

is finitely prevalent in the $P_V(x^*)$. This is demonstrated by showing (1) that the set of function profiles in $P_V(x^*)$ in which $p_i(x^*) \notin Int(\Delta(J))$ for some $i \in N$ is finitely shy in $P_V(x^*)$, which implies that the set of function profiles in $P_V(x^*)$ such that there exists a voter $i$ and a candidate $j$ for which

$$\delta_i^j(p_{-i}(x^*)) = 0$$

is finitely shy in $P_V(x^*)$, and (2) that the set of function profiles in $P_V(x^*)$ such that, for all voters $i$ and candidates $j$, it is the case that

$$Dp_i^j(x^*) = 0,$$

is finitely shy in $P_V(x^*)$.

**Lemma 2** *Choose any point $x^* \in Int(Y)$ and define*

$$B(x^*) = \{p \in P_V(x^*) : \exists j \in J, i \in N \textit{ such that } p_i^j(x^*) = 0\}.$$

*The set $B(x^*)$ is finitely shy in $P_V(x^*)$.*

[9]The analysis would be much more complicated if boundary policy profiles were considered. I conjecture that the results stated here would still hold, however, so long as the policy space is convex, since any policy on the boundary that maximizes an objective function must satisfy the first and second order conditions relative to the interior of the policy space.

*Proof*: Note that all closed sets are completely metrizable in the relative topology induced from the topology of $C^2$ uniform convergence on $P(Y)$ (Aliprantis and Border (1994), p.73). It can be shown that $P_V(x^*)$ is a closed and convex subset of $P(Y)$ and, hence, completely metrizable in the topology of $C^2$ uniform convergence on $P(Y)$. Similarly, it may be verified that $B(x^*)$ is closed and therefore a Borel subset in the topology of $C^2$ uniform convergence.

Consider the following function, which is constant with respect to $Y$:

$$p(\cdot|\alpha) = (\alpha, (1-\alpha)/(|J|-1), \ldots, (1-\alpha)/(|J|-1)),$$

and let $h(\cdot|\alpha) = (p(\cdot|\alpha), \ldots, p(\cdot|\alpha))$ denote a $n$-dimensional profile of identical response functions. Define $H$ as the following one dimensional subspace of $P(Y)$: $H = \{h(\cdot|\alpha)|\alpha \in \mathbb{R}\}$. Since $0 \leq \alpha \leq 1$ implies that $h(\cdot|\alpha) \in P_V(x^*)$, it follows that $\lambda_H(P_V(x^*)) > 0$. We now show that $\lambda_H(B(x^*) + g) = 0$ for any $g \in \mathcal{C}^2$. Consider any $a, b \in \mathbb{R}$ and $s, t \in B(x^*)$ such that

$$\begin{aligned} s + g &= h(\cdot|a) \\ t + g &= h(\cdot|b). \end{aligned}$$

It must be the case then $Dh(\cdot|a) = Ds + Dg$ and $Dh(\cdot|b) = Dt + Dg$. Since $Dh(x|a) = Dh(x|b) = 0$ for all $x \in Y$ and any real numbers $a$ and $b$, it follows that $Ds(x) = -Dg(x)$ and $Dt(x) = -Dg(x)$ for all $x \in Y$, so that $Ds = Dt$. Therefore, if $s_i^j(x^*) = 0$ and $t_i^j(x^*) = 0$ for some $i \in N$ and $j \in J$, then it must be the case that $h(\cdot|a) = g(x^*) = h(\cdot|b)$, which implies that $a = b$.[10]

Fixing $g \in \mathcal{C}^2$, it follows that for each pair $(i, k)$, with $i \in N$ and $k \in J$, there is at most one real number $a$ and one function $s \in B(x^*)$ such that $s_i^k(x^*) = 0$ and $s + g = h(\cdot|a)$. There are at most $|J|n$ such pairs for any given $g \in \mathcal{C}^2$. In other words, for any $g \in \mathcal{C}^2$, $(B + g) \cap H$ contains at most $|J|n$ elements. Since the Lebesgue measure of any finite set is zero, we have that $\lambda_H(B + g) = 0$, so that $B$ is finitely shy relative to $P_V(x^*)$, as was to be shown. ∎

**Lemma 3** *Choose any point $x^* \in Int(Y)$ and define*

$$Z(x^*) = \{p \in P_V(x^*) : \forall j \in J, \forall i \in N, Dp_i^j(x^*) = 0\}.$$

*The set $Z(x^*)$ is finitely shy in $P_V(x^*)$.*

*Proof*: It has already been shown that $P_V(x^*)$ is a completely metrizable convex subset of $P(Y)$. To see that $Z(x^*)$ is a Borel subset in the topology of $C^2$ uniform convergence, note that $Z$ is closed.

Choose a function $f : X \to (0, 1)$ such that $f$ is twice continuously differentiable and, for

[10] In particular, if $s_i^j(x^*) = t_i^j(x^*)$ for $j = 1$, then $a = g_i^1(x^*) = b$. If $s_i^j(x^*) = t_i^j(x^*)$ for $j \neq 1$, then $a = 1 - (|J| - 1)g_i^1(x^*) = b$.

all $x \in X$, $D_x f(x) \neq 0$. Then define

$$\begin{aligned}
p_1(x|\phi) &= \left(\phi f(x_1), \phi(1-f(x_1)), \frac{1-\phi}{|J|-2}, \ldots, \frac{1-\phi}{|J|-2}\right), \\
p_2(x|\phi) &= \left(\phi(1-f(x_1)), \phi f(x_1), \frac{1-\phi}{|J|-2}, \ldots, \frac{1-\phi}{|J|-2}\right), \\
p_3(x) &= \left(\frac{1}{|J|}, \ldots, \frac{1}{|J|}\right), \text{ and} \\
h(\cdot|\phi) &= \left(p_1(\cdot|\phi), p_2(\cdot|\phi), p_3(\cdot), \ldots, p_3(\cdot)\right),
\end{aligned}$$

with $\phi \in \mathbb{R}$.

The set $H = \{h(\cdot|\phi)|\phi \in \mathbb{R}\}$ is a one-dimensional subspace of $P(Y)$. Consider any $\phi$ in the open interval $(0,1)$. By construction, $p_1(\cdot|\phi), p_2(\cdot|\phi)$, and $p_3(\cdot)$ are twice continuously differentiable response functions. Furthermore, $V_k(x)$ is constant for each candidate $k \in J$ and all policy profiles $x \in Y$.[11] From these facts it follows that $\lambda_H(P_V(x^*)) > 0$. It is now shown that $\lambda_H(Z(x^*) + g) = 0$ for any $g \in \mathcal{C}^2$. Suppose that, for some $g \in \mathcal{C}^2$, $(Z(x^*) + g) \cap H$ contains more than one element. Then it must be the case that there exist distinct scalars $a, b \in \mathbb{R}$ and distinct vectors of response functions $s, t \in Z(x^*)$ such that

$$\begin{aligned}
s + g &= h(\cdot|a), \\
t + g &= h(\cdot|b).
\end{aligned}$$

This would imply that $Ds(x) = Dh(x|a) - Dg(x)$ and $Dt(x) = Dh(x|a) - Dg(x)$ for any $x \in Y$. By definition, $s, t \in Z(x^*)$ implies that $Ds(x^*) = Dt(x^*) = 0$, so that

$$Dh(x^*|a) = Dg(x^*) = Dh(x^*|b).$$

In particular, considering the strategy of candidate 1, it must be the case that

$$D_{x_1} h(x^*|a) = D_{x_1} g(x^*) = D_{x_1} h(x^*|b).$$

Where, since $h(x^*|\cdot)$ and $g(x^*)$ are $n \times |J|$ matrices, the differentiation denoted by $D_{x_1} h(x^*|a)$, $D_{x_1} h(x^*|b)$, and $D_{x_1} g(x^*)$ is performed component-wise in each case. Accordingly, this differentiation results in the following:

$$\begin{aligned}
D_{x_1} h(x^*|a) &= (aDf(x^*), -aDf(x^*), 0, \ldots, 0) \text{ and} \\
D_{x_1} h(x^*|b) &= (bDf(x^*), -bDf(x^*), 0, \ldots, 0),
\end{aligned}$$

with $Df(x^*) \neq 0$. It follows then that $Dh(x^*|a) = Dh(x^*|b)$ implies $a = b$, contradicting the supposition that $a$ and $b$ are distinct. Therefore, since $(Z(x^*) + g) \cap H$ contains at most one element, it must be the case that $\lambda_H(Z(x^*) + g) = 0$ for all $g \in \mathcal{C}^2$. Hence, $Z(x^*)$ is finitely shy relative to $P_V(x^*)$, as was to be shown. ∎

[11] Specifically, $V_k(x) = n/|J|$ for all candidates $k$ and all policy profiles $x$.

The next lemma establishes that a finitely prevalent subset of the $n$-dimensional profiles of twice continuously response functions for which $x^*$ maximizes expected vote share is characterized by all voters having different pivot probabilities for any given candidate in $J$.

Before proceeding to formally stating and proving the lemma, it is illustrative to describe the logic of the proof. The first recognition is that it is sufficient to consider any pair of voters (say, voters 1 and 2) and any candidate (say, candidate 1) and show that the set of profiles of response functions that lead to equal pivot probabilities for those two voters for that candidate is a finitely shy subset of $P(x^*)$. The set of profiles of response functions such that, for any candidate, the pivot probabilities for that candidate for more than one pair of voters are equal is a subset of the set of response functions at which at least one pair of voters have equal pivot probabilities for some candidate. Since the numbers of voters and candidates are each finite and the union of finitely many finitely shy sets is itself finitely shy, this approach is sufficient to show that the result holds.

The second fact motivating the proof of the result is that the pivot probability for voter 1 (for example) is a function of all other voters' behaviors at $x^*$ (i.e., $p_{-1}(x^*)$) and not his or her own behavior (i.e., $p_1(x^*)$). In addition, this probability is a function only of the value of all other voters' response functions at $x^*$. This greatly simplifies the problem in the sense that one can deal only with constant response functions (or, in other words, one can identify each response function with a unique vector in $\Delta(J)$). Using these facts, the proof essentially holds the response functions of voters $3, 4, \dots, n$ constant (after translation by $g \in \mathcal{C}^2$) and then considers whether $p_1(x^*)$ restricts the set of $p_2(x^*)$ such that $\delta_1^j(p_{-1}(x^*)) = \delta_2^j(p_{-2}(x^*))$ for some candidate $j$ to a subspace with empty interior relative to $\Delta(J)$. If this is the case, then the lemma follows.

Broadly speaking, the proof consists of four steps. The first step, after constructing a subspace of constant response functions, is the expression of voter 1's pivot probability for candidate 1 as a linear function of voter 2's response function, holding the response functions of the other $n-2$ voters constant. The logic of this step is that, in most cases (in terms of the other $n-2$ voters' response functions), any perturbation of voter 2's response function will result in a different pivot probability for candidate 1. The second step of the proof is demonstrating that this is indeed the case. The third step of the proof deals with situations in which perturbing voter 2's response function will not alter voter 1's pivot probability. These cases are rare, but important. This case is dealt with by considering voter 3 and perturbing his or her behavior.[12] This step is slightly complicated by the fact that then the $n-3$ remaining voters' behaviors are held fixed. If these $n-3$ response functions match up in a very specific way (which can not be ruled out), then we must go further, considering voters 4,5, and so on. The final step of the proof is showing that this process need include no more than the smallest strict majority of the voters. At this point, it is impossible for the response functions of the other $(n-3)/2$ voters[13]

[12] An example of such a situation with three voters and two candidates is when voter 3 votes for candidate 1 with probability 1/2 and candidate two with probability 1/2. In this case, voter 1's pivot probability for either candidate is 1/2, regardless of voter 2's behavior. If voter 3's behavior is perturbed slightly, then this is no longer the case. I thank a referee for suggesting this example.

[13] Or, $n/2-1$ voters if $n$ is even.

to match up so that perturbing the $(n+1)/2th$ voter's[14] response function does not affect voter 1's pivot probability for candidate 1.

The proof, while complicated in some ways, has a fairly straightforward logic behind it. Any voter's pivot probability for a given candidate is simply a sum of the product of the other voters' response functions over a subset of the possible vote profiles (namely, the vote profiles in which that voter's vote for the candidate in question is decisive). Lemma 2 allows us to consider only cases in which all of these response functions are in the strict interior of the $|J|-1$ dimensional simplex. This turns out to guarantee that varying one of the voters' response functions will generally change this sum of products. The complicated steps involve ensuring that the special cases where this is not the case are nongeneric.

**Lemma 4** *For any point $x^* \in Int(Y)$, the set*

$$T(x^*) = \{p \in P_V(x^*) \setminus B(x^*) : \exists i \in N, k \in N \setminus \{i\}, j \in J, \delta_i^j(p_{-i}(x^*)) = \delta_k^j(p_{-k}(x^*))\}$$

*is finitely shy relative to $P_V(x^*)$.*

*Proof*: It has been demonstrated previously that $P_V(x^*)$ is a completely metrizable convex subset of $P(Y)$. To see that $T(x^*)$ is a Borel subset, note that it is a closed set intersected with the complement of a Borel set (since $B(x^*)$ is a Borel set).

Let $h(\cdot|\alpha_1,\ldots,\alpha_n) = (h_1(\cdot|\alpha_1), h_2(\cdot|\alpha_2), \ldots, h_n(\cdot|\alpha_n)) = (\alpha_1,\ldots,\alpha_n)$, for $\alpha_1,\ldots,\alpha_n \in \mathbb{R}^{|J|}$ (in other words, each voter's response function is a constant function on $X$). Let $O(J) = \{\alpha \in \mathbb{R}^J : \sum_{j=1}^{|J|} \alpha^j = 1\}$ and denote by $H$ the $n(|J|-1)$-dimensional subset of $\mathcal{C}^2$ generated by $h$:

$$H = \{h(\cdot|\alpha_1,\ldots,\alpha_n) : \alpha_i \in O(J); \forall i \in N\}.$$

Let $\alpha = (\alpha_1,\ldots,\alpha_n)$. It is clear that $\lambda_H(P_V(x^*)) > 0$ since $h(\cdot|\alpha) \in P_V(x^*)$ for all $\alpha$ such that $\alpha_i^j \geq 0$, for all $i \in N$ and all $j \in J$.

Let $h_{-1}(x^*|a_{-1}) - g_{-1}(x^*)$ denote the vector of response functions $(h_2(\cdot|a_2) - g_2, h_3(\cdot|a_3) - g_3, \ldots, h_n(\cdot|a_n) - g_2)$, evaluated at $x^*$ and similarly for $h_{-2}(x^*|a_{-2}) - g_{-2}(x^*)$, $h_{-3}(x^*|a_{-3}) - g_{-3}(x^*)$, and so forth. Fix $g \in \mathcal{C}^2$ and define

$$\begin{aligned}\mathcal{A}_g(i,j,k) = \ & \{a \in (O(J))^n : \\ & \delta_i^k(h_{-i}(x^*|a_{-i}) - g_{-i}(x^*)) = \delta_j^k(h_{-j}(x^*|a_{-j}) - g_{-j}(x^*))\}.\end{aligned}$$

$\mathcal{A}_g(i,j,k)$ is the set of $a \in (O(J))^n$ such that

$$s + g = h(\cdot|a)$$

for some $s \in T(x^*)$. Accordingly, if $\lambda_H(\mathcal{A}_g(i,j,k)) = 0$ for arbitrary $g \in \mathcal{C}^2$, $i \neq j \in N$, and $k \in J$, it follows that $T(x^*)$ is finitely shy in $P_V(x^*)$.

I now consider voters 1 and 2 and candidate 1 (without loss of generality) and derive voter 1's pivot probability as a function of voter 2's behavior (*i.e.*, $a_2$), holding the behavior of the remaining voters (*i.e.*, $a_3,\ldots,a_n$) constant.

[14] Or, the $n/2+1th$ voter if $n$ is even.

Given $g \in \mathcal{C}^2$, suppose that $s \in P_V(x^*)$, with $s = h(\cdot|a) - g$ for some $a \in O(J)^n$, and that voter 1 and voter 2 have equal pivot probabilities for candidate 1:

$$\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*)) = \delta_2^1(h_{-2}(x^*|a_{-2}) - g_{-2}(x^*)).$$

Now express $\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*))$ as a function of $h_2 - g_2$ as follows:

$$\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*)) = \sum_{j \in J}(h_2^j(x^*|a_2) - g_2^j(x^*))K_{1,2}^1(j, h - g),$$

where $K_{i,l}^k(j, q)$ is the probability that voter $i$ is pivotal for candidate $k$, conditional on voter $l$ voting for candidate $j$ and the $n$-dimensional profile of response functions $q$.[15] Substituting $h_2^j(x^*|a_2) = a_2^j$, this becomes

$$\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*)) = \sum_{j \in J}(a_2^j - g_2^j(x^*))K_{1,2}^1(j, h - g).$$

Note that voter 2's pivot probability for candidate 1, $\delta_2^1$, is not a function of voter 2's behavior, $h_2 - g_2$. By supposition, $\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*)) = \delta_2^1(h_{-2}(x^*|a_{-2}) - g_{-2}(x^*))$. Therefore, we need to show that

$$\delta_2^1(h_{-2}(x^*|a_{-2}) - g_{-2}(x^*)) = \sum_{j \in J}(a_2^j - g_2^j(x^*))K_{1,2}^1(j, h - g). \tag{2}$$

holds for a subset of $O(J)$ possessing Lebesgue measure zero.

There are two cases to consider. The first case (Case I) is if there exists two candidates $j, k \in J$ such that $K_{1,2}^1(j, h - g) \neq K_{1,2}^1(k, h - g)$. This case holds "most" of the time. The second case (Case II) is when, for all pairs of candidates $j, k \in J$, we have that $K_{1,2}^1(j, h - g) = K_{1,2}^1(k, h - g)$. I deal with the cases in order. Since $K_{1,2}^1(j, h - g)$ is a function of $a_3, \ldots, a_n$, these two cases correspond to different configurations of behavior by the remaining $n - 2$ voters.

Case I: *There exist two candidates* $j, k \in J$ *such that* $K_{1,2}^1(j, h - g) \neq K_{1,2}^1(k, h - g)$.

In this case, the set of $a_2$ that satisfy Equation 2 possess dimension no greater than $|J| - 2$, which is strictly less than the dimensionality of $O(J)$ (which is $|J| - 1$), implying that this subset possesses Lebesgue measure zero in $O(J)$. The Cartesian product of this subset and $O(J)^{n-1}$ lies within $O(J)^n$. Since the subset defined by Equation 2 has measure zero, Fubini's theorem [Halmos, (1974), Theorem A, p. 147], then implies the set of $a \in \mathcal{A}_g(1, 2, 1)$ such that Case I holds, defined as

$$\overline{\mathcal{A}}_g(1, 2, 1) = \{a \in \mathcal{A}_g(1, 2, 1) : \exists j, k \in J \text{ s.t. } K_{1,2}^1(j, h - g) \neq K_{1,2}^1(k, h - g),$$

[15] While $K_{1,2}^1(j, h - g)$ is conditional on the *action* of voter 2, the construction of $K_{i,l}^k$ implicitly includes the response functions of the $n-2$ voters other than 1 and 2 (*i.e.*, $a_3, \ldots, a_n$). By holding $K_{i,l}^k$ fixed, we are supposing that these $n - 2$ response functions are held fixed. Below, we define versions of $K$ that are conditioned on the actions of more voters (*i.e.*, voters 3, 4, and so on). The logic of those conditional probabilities is analogous to that of $K_{1,2}^1(j, h - g)$.

possesses Lebesgue measure zero in $O(J)^n$.

Case II: *For all pairs of candidates* $j, k \in J$, $K^1_{1,2}(j, h-g) = K^1_{1,2}(k, h-g)$ *holds.*

In this second case, voter 2's behavior (*i.e.,* $a_2$) does not affect voter 1's pivot probability for candidate 1.[16] Therefore, I now consider voter 3 and expand Equation 2 to include voter 3's behavior, obtaining the following:

$$\delta^1_2(h_{-2}(x^*|a_{-2}) - g_{-2}(x^*)) = \sum_{j \in J}(a^j_2 - g^j_2(x^*)) \sum_{j^3 \in J}(a^{j^3}_3 - g^{j^3}_3(x^*))K^1_{1,2,3}(j, j^3, h-g),$$

where $K^1_{1,2,3}(j, j^3, h-g)$ is defined in a manner analogous to $K^1_{1,2}(j, h-g)$, above: it is the probability that voter 1 is pivotal for candidate 1, conditional on voter 2 voting for candidate $j$, voter 3 voting for candidate $j^3$, and the $n$-dimensional profiles of response functions $h-g$. Now consider varying $a_3$. If there exists some $j, k, j^3 \in J$ such that

$$K^1_{1,2,3}(j, j^3, h-g) \neq K^1_{1,2,3}(k, j^3, h-g), \tag{3}$$

then the set of $a_3$ for which case (2) holds possesses Lebesgue measure zero in $O(J)$. To see this, first note that

$$\begin{aligned}
&K^1_{1,2}(j, h-g) - K^1_{1,2}(k, h-g) \\
&= \textstyle\sum_{j^3 \in J}(a^{j^3}_3 - g^{j^3}_3(x^*))K^1_{1,2,3}(j, j^3, h-g) - \sum_{j^3 \in J}(a^{j^3}_3 - g^{j^3}_3(x^*))K^1_{1,2,3}(k, j^3, h-g) \\
&= \textstyle\sum_{j^3 \in J}(a^{j^3}_3 - g_3(x^*))[K^1_{1,2,3}(j, j^3, h-g) - K^1_{1,2,3}(k, j^3, h-g)].
\end{aligned}$$

Then, supposing that $K^1_{1,2,3}(1, q, h-g) \neq K^1_{1,2,3}(2, q, h-g)$ for some $q \in J$, $K^1_{1,2}(j, h-g) - K^1_{1,2}(k, h-g) = 0$ implies that (leaving the $h-g$ argument implicit for reasons of space)

$$0 = \sum_{j^3 \in J}(a^{j^3}_3 - g_3(x^*))[K^1_{1,2,3}(1, j^3) - K^1_{1,2,3}(2, j^3)], \tag{4}$$

Since $K^1_{1,2,3}(1, q, h-g) - K^1_{1,2,3}(2, q, h-g)] \neq 0$, this implies that (holding $a_1, a_2, a_4, \ldots, a_n$ constant) the set of $a_3$ that solves Equation 4 is of dimensionality no greater than $|J| - 2$. This fact plus Fubini's theorem implies that, the set of solutions in $O(J)^n$ to Equation 4 must possess dimensionality no greater than $n(|J| - 1) - 1$, which implies that the Lebesgue measure (in $O(J)^n$, which is of dimensionality $n(|J| - 1)$ of this set must be zero.

To finish this step of the proof, suppose that Equation 3 does not hold for any $j, j^3, k \in J$. The above argument for voter 3 can be applied iteratively, removing (i.e., conditioning upon the actions of) additional voters one at a time and checking a condition analogous to Equation 3. Specifically, if we remove voters as ordered by their subscript,[17] and are considering voter $l > 3$, the analogue to Equation 3 is

$$K^1_{1,2,\ldots,l}(j, j^3, \ldots, j^{l-1}, j^l, h-g) \neq K^1_{1,2,\ldots,l}(j, j^3, \ldots, \hat{j}^{l-1}, j^l, h-g) \tag{5}$$

[16] Similarly, voter 1's behavior (*i.e.,* $a_1$) does not affect voter 2's pivot probability for candidate 1.

[17] This choice of order is unnecessary, but convenient.

for some $j^{l-1}, \hat{j}^{l-1}, j^l \in J$.[18] If, at any voter $l$, Equation 5 holds, then the set of $a_l$ such that $K^1_{1,2}(j, h-g) = K^1_{1,2}(k, h-g)$ for all $j, k \in J$ possesses Lebesgue measure zero in $O(J)$. (The process of proving this involves an extended version of the argument derived following Equation 4, above.)

Now let $l = (n+3)/2$ (or $n/2+2$ if $n$ is even). In this case, it turns out that Equation 5 *must* be satisfied. To see this, consider the case where $n$ is odd[19] and $j = j^3 = \ldots = j^{l-1} = j^l = 1$. In this case, the probability of voter 1 being pivotal for candidate 1, conditional upon voters $2, 3, \ldots, (n+3)/2$ (*i.e.*, a strict majority of the voters) voting for candidate 1 is 0, as voter 1's vote choice can not affect the outcome of the election. If, on the other hand, voter $l$ votes for (say) candidate 2 (*i.e.*, $j^l = 2$), then voter 1's pivot probability for candidate 1 is positive by the supposition that no voter's response function assigns any candidate zero probability (*i.e.*, $h - g \in T(x^*) \Rightarrow h - g \notin B(x^*)$).

Writing this formally, it is the case that

$$K^1_{j,j^3,\ldots,j^l,1} = \begin{cases} \kappa > 0 & \text{if } j^l \neq 1 \\ 0 & \text{otherwise.} \end{cases}$$

For our purposes, we do not need to know the exact value of $\kappa$ (which may depend upon the value of $j^l$).[20] Our sole interest in $\kappa$ is that it is strictly greater than zero for any $j^l \neq 1$.

I now claim that the set of $a_l$ such that Equation 5 does not hold must possess Lebesgue measure zero in $O(J)$. To see this, note that

$$\begin{aligned} K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=1,h-g) &= \sum_{j^l \in J} (a_l^{j^l} - g_l^{j^l}(x^*)) K^1_{1,2,\ldots,l}(1,1\ldots,1,j^l,h-g) \\ &= \sum_{j^l \in J \setminus \{1\}} (a_l^{j^l} - g_l^{j^l}(x^*)) K^1_{1,2,\ldots,l}(j,j^2\ldots,j^l,h-g) \end{aligned}$$

(this step follows because $K^1_{j,j^3,\ldots,j^{l-1},j^l=1,1} = 0$) and that

$$K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=2,h-g) = \sum_{j^l \in J} (a_l^{j^l} - g_l^{j^l}(x^*)) K^1_{1,2,\ldots,l}(1,1\ldots,2,j^l,h-g),$$

so that

$$\begin{aligned} &K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=1,h-g) - K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=2,h-g) \\ &= \sum_{j^l \in J} \left[ (a_l^{j^l} - g_l^{j^l}(x^*)) - (a_l^{j^l} - g_l^{j^l}(x^*)) \right] K^1_{1,2,\ldots,l}(1,1\ldots,1,j^l,h-g). \end{aligned}$$

[18]Note that the order of subscripts does not matter: one could, for example, phrase this condition as

$$K^1_{1,2,\ldots,l}(j,j^3,\ldots,j^{l-1},j^l,h-g) \neq K^1_{1,2,\ldots,l}(j,\hat{j}^3,\ldots,j^{l-1},j^l,h-g) \quad (6)$$

for some $j^3, \hat{j}^3, j^l \in J$. This is because the simple plurality rule considered here is anonymous.

[19]The case where $n$ is even is analogous.

[20]It is easily derived though: the actual value of $K^1_{j,j^3,\ldots,j^{l-1},1}$ for $j^{l-1} \neq 1$ is 0.5 multiplied by the probability of all $n/2 - 1$ remaining voters voting for $j^{l-1}$ if $n$ is even. If $n$ is odd, then it is 0.5 multiplied by the probability of all $(n-1)/2$ remaining voters voting for some candidate other than $j^{l-1}$ (including candidate 1). While these might obviously be very small numbers, they are not zero, by the requirement that $h - g \notin B(x^*)$.

Then, letting

$$\hat{K}(j^l) = K^1_{1,2,\ldots,l}(1,1,\ldots,1,j^l,h-g) - K^1_{1,2,\ldots,l}(1,1,\ldots,2,j^l,h-g),$$

it follows that

$$K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=1,h-g) \quad = \quad K^1_{1,2,\ldots,l-1}(1,1,\ldots,j^{l-1}=2,h-g)$$

holds only if,

$$\begin{aligned} \sum_{j^l \in J}(a_l^{j^l} - g_l^{j^l}(x^*))\hat{K}(j^l) &= 0 \\ \sum_{j^l \in J\setminus\{1\}} \frac{(a_l^{j^l} - g_l^{j^l}(x^*))\hat{K}(j^l)}{K^1_{1,2,\ldots,l}(1,1\ldots,2,1,h-g)} + g_l^1(x^*) &= a_l^1 \end{aligned} \qquad (7)$$

and, since $h-g \notin B(x^*)$, it follows that $K^1_{1,2,\ldots,l}(1,2,\ldots,2,j^l) > 0$ for all $j^l$. Since $a_l^1$ is determined uniquely in Equation 7, the set of $a_l$ satisfying Equation 7 must possess Lebesgue measure zero in $O(J)$.[21] Thus, the set of $a_1,\ldots,a_n$ such that $K^1_{1,2}(j,h-g) = K^1_{1,2}(k,h-g)$ possesses Lebesgue measure zero in $O(J)^n$.

Letting

$$\underline{\mathcal{A}}_g(1,2,1) = \mathcal{A}_g(1,2,1) \setminus \overline{\mathcal{A}}_g(1,2,1),$$

denote the subset of $\mathcal{A}(1,2,1)$ in which Case II holds, it follows that $\underline{\mathcal{A}}_g(1,2,1)$ possesses Lebesgue measure zero in $O(J)^n$, further implying (once again by Fubini's theorem) that $\underline{\mathcal{A}}_g(1,2,1)$ possesses Lebesgue measure zero in $O(J)^n$.

To conclude the proof, first note that the Lebesgue measure of $\mathcal{A}_g(1,2,1)$ in $O(J)^n$ is less than or equal to the sum of its Lebesgue measure in Cases I and II:

$$\lambda_H(\mathcal{A}_g(1,2,1)) \leq \lambda_H(\overline{\mathcal{A}}_g(1,2,1)) + \lambda_H(\underline{\mathcal{A}}_g(1,2,1)).$$

Thus, the Lebesgue measure of $\mathcal{A}_g(1,2,1)$ in $O(J)^n$ must be zero. Finally, note that the choice of candidates and voters is arbitrary, thus proving the result for $\mathcal{A}_g(i,j,k)$, $i,j \in N$, and $k \in J$. Hence, $T(x^*)$ is finitely shy relative to $P_V(x^*)$, as was to be shown. ∎

The final lemma states that, given any point $x^* \in Int(Y)$, the set of profiles of response functions $p \in (P_V(x^*) \setminus (B(x^*) \cup Z(x^*) \cup T(x^*)))$ that simultaneously satisfy, for each candidate $k \in J$, the necessary first and second order conditions for maximization of expected vote share and the necessary first order conditions for maximization of probability of victory at $x^*$ is finitely shy with respect to the set of profiles of response functions that satisfy, for each candidate $k \in J$, the necessary first and second order conditions for expected vote share maximization. This result is used to prove the paper's main results, which state that the set

[21]The process described here, more generally, can be thought of as rewriting $\delta_1^1(h_{-1}(x^*|a_{-1}) - g_{-1}(x^*))$ as a function of a $|J| \times |J| \times \ldots \times |J|$ "hypermatrix." Each voter reduces the dimensionality of this hypermatrix. Hopefully the derivation in terms of sums makes the logic more transparent.

of profiles of response functions which simultaneously satisfy the necessary first and second order conditions for maximization of both objective functions is finitely shy with respect to the set of profiles that satisfy the first and second order conditions under expected vote share maximization.

**Lemma 5** *For any point $x^* \in Int(Y)$, the set*

$$R1(x^*) = \{p \in P_V(x^*) \setminus (Z(x^*) \cup B(x^*) \cup T(x^*)) : DR_j(x^*) = 0 \; \forall j \in J\}$$

*is finitely shy relative to $P_V(x^*)$.*

*Proof*: That $P_V(x^*)$ is a completely metrizable convex subset of $P(Y)$ has been demonstrated previously. It is easily verified that $R1(x^*)$ is a Borel subset (it is the intersection of a Borel set with a closed set).

Let $f(\cdot|\alpha) : X \times \mathbb{R} \to (-1/(2|J|), 1/(2|J|))$ be a twice continuously differentiable function with $f(x_1^*) = 0$ and $D_x f(x_1^*) = 1$..[22] Define

$$h_i(y|\alpha_i, \beta_i) = (\alpha_i f(x_1) + \beta_i, 1/|J| - \alpha_i f(x_1) - \beta_i, 1/|J|, \ldots, 1/|J|)$$

for all $i \in N$. Let

$$\begin{aligned} \alpha &= (\alpha_1, \ldots, \alpha_n), \\ \beta &= (\beta_1, \ldots, \beta_n), \\ h(\cdot|\alpha, \beta) &= (h_1(\cdot|\alpha_1, \beta_1), \ldots, h_n(\cdot|\alpha_n, \beta_n)), \end{aligned}$$

and let $H = \{h(\cdot|\alpha, \beta) : \sum_{i=1}^n \alpha_i = 0, \beta \in \mathbb{R}^n\}$. This is a $2n - 1$ dimensional subspace of $\mathcal{C}^2$.

Note that, for any $g \in \mathcal{C}^2$, any voter $i \in N$, any candidate $j \in J$, and any $\beta \in \mathbb{R}^n$, the following holds for all $\alpha, \alpha' \in \mathbb{R}^n$:

$$\delta_i^j(g_{-i}(x^*) + h_{-i}(x^*|\alpha, \beta)) = \delta_i^j(g_{-i}(x^*) + h_{-i}(x^*|\alpha', \beta)),$$

(where the subscript $-i$ denotes the appropriate vector of functions for all $j \in N \setminus \{i\}$). In other words, a fixed value of $\beta$ "pins down" the voters' pivot probabilities. Similarly, for any $g \in \mathcal{C}^2$, any voter $i \in N$, any candidate $j \in J$, and any $\alpha$, the following holds for all $\beta, \beta' \in \mathbb{R}^n$:

$$D(g_{-i}(x^*) + h_{-i}(x^*|\alpha, \beta)) = D(g_{-i}(x^*) + h_{-i}(x^*|\alpha, \beta')),$$

so that a fixed value of $\alpha$ pins down the gradients of voters' behaviors.

Note that $\lambda_H(P_V(x^*)) > 0$ since $h(\cdot|\alpha, \beta) \in P_V(x^*)$ if $\sum_{i=1}^n \alpha_i = 0$ and, for all $i \in N$, $\beta_i \in (-1/(2\|J|), 1/2|J|)$. It is now shown that $\lambda_H(R1(x^*) + g) = 0$ for any $g \in \mathcal{C}^2$. To prove

[22] The notation $x_1^*$ denotes candidate 1's position in policy profile $x^*$. The function $f$ depends only on candidate 1's policy position.

this, it suffices to show that, for arbitrary fixed $g \in \mathcal{C}^2$ and for all $\beta$ such that $\delta_1^1(g_{-1}(x^*) + h_{-1}(x^*|\cdot,\beta)) \neq \delta_2^1(g_{-2}(x^*) + h_{-2}(x^*|\cdot,\beta))$,[23] the set

$$S_g(\beta) = \{\alpha \in \mathbb{R}^n : g + h(x^*|\alpha,\beta) \in R1(x^*)\}$$

possesses Lebesgue measure zero in $\mathbb{R}^n$.

To see why this is sufficient, fix $\alpha, \beta$ and let $s = g + h(\cdot|\alpha,\beta)$. Then note that $s \in R1(x^*)$ implies that there exists a distinct pair of voters $i, k$ and a candidate $j$ such that

$$D_{x_j} s_i(x^*) > 0 > D_{x_j} s_k(x^*).$$

Therefore, one can examine voters 1 and 2 and candidate 1 without loss of generality. Second, note that $s \in R1(x^*)$ implies that $\delta_i^j(s_{-i}(x^*)) > 0$ for all $i \in N$ and $j \in J$. Finally, $\beta$ such that $\delta_1^1(s_{-1}(x^*)) = \delta_2^1(s_{-2}(x^*))$ implies that $s \in T(x^*)$ and hence $s \notin R1(x^*)$.

Noting that $s = g + h(\cdot|\alpha,\beta) \in R1(x^*)$ implies that

$$\sum_{i=1}^{n} \delta_i(s_{-i}(x^*))[D_{x_1} g_i(x^*) + D_{x_1} h_i(x^*|\alpha_i,\beta_i)] = 0,$$

it follows that, letting $K(\alpha,\beta) = -\sum_{i=3}^{n} \delta_i(s_{-i}(x^*))[D_{x_1} g_i(x^*) + D_{x_1} h_i(x^*|\alpha_i,\beta_i)]$[24],

$$\begin{array}{c} \delta_1(s_{-1}(x^*))[D_{x_1} g_1(x^*) + D_{x_1} h_1(x^*|\alpha_1,\beta_1)] \\ +\delta_2(s_{-2}(x^*))[D_{x_1} g_2(x^*) + D_{x_1} h_2(x^*|\alpha_2,\beta_2)] \end{array} = K(\alpha,\beta).$$

Substituting for $h_1$ and $h_2$,

$$\delta_1(s_{-1}(x^*))[D_{x_1} g_1(x^*) + \alpha_1] + \delta_2(s_{-2}(x^*))[D_{x_1} g_2(x^*) + \alpha_2] = K(\alpha,\beta).$$

A sufficient condition for $\lambda_H(R1(x^*) + g) = 0$ is, for any fixed $\alpha_{-2} = \tilde{\alpha}_{-2}$, there exists a unique value of $\alpha_2$ such that $s, t \in R1(x^*)$, $s = g + h(\alpha_2, \tilde{\alpha}_{-2}, \beta)$, and $t + g = h(\alpha_2', \tilde{\alpha}_{-2}, \beta)$ jointly imply that $\alpha_2 = \alpha_2'$. In other words, a necessary condition for $\lambda_H(R1(x^*) + g) > 0$ is that there exist some $\alpha_{-2}, \beta$ such that

$$\begin{aligned} K(\alpha,\beta) &= \delta_1(s_{-1}(x^*))[D_{x_1} g_1(x^*) + \alpha_1] + \delta_2(s_{-2}(x^*))[D_{x_1} g_2(x^*) + z] \\ z\delta_2(s_{-2}(x^*)) &= K(\alpha,\beta) - \delta_1(s_{-1}(x^*))[D_{x_1} g_1(x^*) + \alpha_1] - \delta_2(s_{-2}(x^*))[D_{x_1} g_2(x^*)] \\ z &= \frac{K(\alpha,\beta) - \delta_1(s_{-1}(x^*))[D_{x_1} g_1(x^*) + \alpha_1] - \delta_2(s_{-2}(x^*))[D_{x_1} g_2(x^*)]}{\delta_2(s_{-2}(x^*))} \qquad (8) \end{aligned}$$

for more than one value of $z$. However, $s \in R1(x^*)$ implies that $\delta_2(s_{-2}(x^*)) > 0$, so that $z$ is uniquely determined by Equation 8. Since $\alpha_{-2}$ and $\beta$ are arbitrary in Equation 8 (except that $\beta$ must, of course, be such that $s \notin T(x^*)$), it must be the case that $\lambda_H(R1(x^*) + g) = 0$ because the dimensionality of the set of solutions to Equation 8 must be no greater than $2n - 2$ (implying that its $2n - 1$ dimensional Lebesgue measure is zero). Thus, $R1(x^*)$ is finitely shy relative to $P_V(x^*)$, as was to be shown. ∎

[23] Recall that specifying the vector $\beta$ is sufficient to generate the the pivot probabilities for all voters $i \in N$ and all candidate $j \in J$, even with $\alpha$ left unspecified.

[24] Note that $K(\alpha,\beta)$ is constant with respect to $\alpha_1$ and $\alpha_2$.

I now prove the following theorem, which states that a policy profile $x^*$ that simultaneously satisfies each candidate's first and second order conditions for maximization of expected vote share generically (in terms of the voters' response functions) does not do so for each candidate's probability of victory as well.

**Theorem 1** *For any point $x^* \in Int(Y)$, the set $P_{V,R}(x^*, J)$ is finitely shy in $P_V(x^*)$.*

*Proof*: Note that $P_{V,R}(x^*, J) \subset R1(x^*) \cup B(x^*) \cup Z(x^*)$. By Lemma 2, $B(x^*)$ is finitely shy in $P_V(x^*)$. By Lemma 3, $Z(x^*)$ is finitely shy in $P_V(x^*)$. By Lemma 5, $R1(x^*)$ is finitely shy in $P_V(x^*)$. Thus, $P_{V,R}(x^*, J)$ is the subset of a finite union of sets that are finitely shy in $P_V(x^*, J)$ and hence finitely shy in $P_V(x^*, J)$ as well. ∎

I now state the main result, which states that a policy profile $x^*$ that satisfies the first and second order conditions for maximization of expected vote share for any candidate $j$ generically does not do so for that candidate's probability of victory. This result is stronger than Theorem 1 in that the other candidates' objectives are left arbitrary.

Before presenting the main result, define the following sets for all candidates $j \in J$ and all interior policies $x^* \in Int(Y)$:[25]

$$
\begin{aligned}
P_V^j(x^*) &= \left\{p \in P(Y) : DV_j(x^*) = 0 \text{ and } D^2V_j(x^*) \text{ is n.s.d,}\right\} \\
P_R^j(x^*) &= \left\{p \in P(Y) : DR_j(x^*) = 0 \text{ and } D^2R_j(x^*) \text{ is n.s.d,}\right\} \\
P_{V,R}^j(x^*) &= P_V^j(x^*) \cap P_R^j(x^*) \\
B^j(x^*) &= \{p \in P_V(x^*) : \exists j \in J, i \in N \text{ such that } p_i^j(x^*) = 0\} \\
Z^j(x^*) &= \{p \in P_V(x^*) : \forall i \in N, Dp_i^j(x^*) = 0\} \\
T^j(x^*) &= \{p \in P_V(x^*) \setminus B(x^*) : \exists i \neq k \in N, \delta_i^j(p_{-i}(x^*)) = \delta_k^j(p_{-k}(x^*))\} \\
R1^j(x^*) &= \{p \in P_V(x^*) \setminus (Z(x^*) \cup B(x^*) \cup T(x^*)) : DR_j(x^*) = 0 \}
\end{aligned}
$$

Note that, for any $j \in J$, the proofs of Lemmas 2, 3, 4, and 5 can be applied to prove that $B^j(x^*)$, $Z^j(x^*)$, $T^j(x^*)$, and $R1^j(x^*)$ are each finitely shy in $P_V^j(x^*)$. Thus, the following result is stated without proof, as it is a mirror of the proof of Theorem 1.

**Theorem 2** *For any $j \in J$, $x^* \in Int(Y)$, $P_{V,R}^j(x^*)$ is finitely shy with respect to $P_V^j(x^*)$.*

Theorem 2 states that, when considering an arbitrary profile of response functions and a vector of opponents' policies under which an interior policy $x^*$ satisfies the necessary conditions for maximization of candidate $j$'s expected vote share, it is generally not the case that the first and second order conditions for maximization of the candidate's probability of victory will be satisfied at $x^*$ as well. One conclusion to be drawn is that, in general, the best

[25]It might be useful to note that $P_V(x^*) = \cap_{j\in J} P_V^j(x^*)$, $P_R(x^*) = \cap_{j\in J} P_R^j(x^*)$, $P_{V,R}(x^*) = \cap_{j\in J} P_{V,R}^j(x^*)$, $B(x^*) = \cap_{j\in J} B^j(x^*)$, $Z(x^*) = \cap_{j\in J} Z^j(x^*)$, $T(x^*) = \cap_{j\in J} T^j(x^*)$, and $R1(x^*) = \cap_{j\in J} R1^j(x^*)$.

response correspondences generated by maximization of probability of victory and maximization of expected vote share maximization will differ. A second conclusion to be drawn is that the genericity found in Theorem 1 does not depend on the assumption that all candidates share the same objective. In other words, *regardless of what the other candidates choose, platforms satisfying the first and second order conditions for maximization of one objective generically do not satisfy the first and second order conditions for the other. Thus, the results do not depend on the assumption that the candidates all share the same objective function.*

# 6 Conclusions

In this paper I have shown that satisfaction of the first and second order conditions for maximization of a candidate's expected vote share generically implies the violation of the first and second order conditions for maximization of that candidate's probability of victory. Making the point another way, the results presented in this paper demonstrate that the predictions of game theoretic models of electoral competition with probabilistic voters will almost always depend upon the assumed functional form of politicians' objectives. Furthermore, this is true for two commonly used versions of "office motivation."

The paper's results hold for any policy on the interior of the policy space as long as voters' behaviors are only restricted to be twice continuously differentiable functions of the policy profile chosen by the candidates. An important implication of this result is that best response equivalence between these two objectives is "almost never" satisfied. This result is in accordance with the tenor of the results of Aranson, Hinich, and Ordeshook (1974), which also show that equivalence between maximization of vote share and maximization of probability of victory is a rare event, though in a different framework.

The importance of these results lies in the research topics which remain open due to the frequent failure of equivalence to hold. In particular, what are the properties of electoral competition under different objective functions? Are equilibrium outcomes under one objective function more representative than under another? What is the relative "punishment" (in terms of decreased chances of victory) of candidates who seek to maximize vote share under different electoral rules?

There are several questions regarding candidates objective functions which remain open. Perhaps the most relevant of these questions is what are the effects of different electoral institutions on equivalence between candidate objective functions? For example, we have not examined the properties of proportional representation, multiple winners, multiple ballot systems (e.g., simple majority rule systems with runoffs or party based systems with primaries), or different scoring rules such as approval voting and the Borda count.

More immediate extensions of the model include the following. It may be of interest to restrict attention to voter response functions which are symmetric. If voter $i$ possesses a symmetric response, then if 2 or more candidates choose the same policy, voter $i$ votes for each such candidate with equal probability (this is a property of logit and probit response functions in a world of policy-motivated voters, for example).[26] Also, I do not examine at least one other plausible objective function: maximization of expected margin of victory. Aranson, Hinich, and Ordeshook (1974), Hinich (1977), Ledyard (1984), and Patty (2001) each examine this objective function, but primarily in the context of 2 candidate contests. Finally, the question of abstention has not been dealt with in this paper. It is conjectured that allowing for abstention will only strengthen the tenor of the results obtained in this paper.[27]

# References


Aliprantis, Charalambos D. and Kim C. Border. 1994. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Berlin: Springer-Verlag.

Anderson, Robert and William R. Zame. 2000. "Genericity with Infinitely Many Parameters," *Advances in Theoretical Economics* 1: 1-62.

Aranson, P. H., M. J. Hinich, and P. C. Ordeshook. 1974. "Election Goals and Strategies: Equivalent and Nonequivalent Candidate Objectives." *American Political Science Review*, 68: 135–52.

Coughlin, P. J. 1992. *Probabilistic Voting Theory*. Cambridge: Cambridge University Press.

Duggan, John. 2000. "Equilibrium Equivalence Under Expected Plurality and Probability of Winning Maximization," *mimeo*, University of Rochester.

Halmos, Paul R. 1974. *Measure Theory*. Springer-Verlag.

Hinich, M. 1977. "Equilibrium in Spatial Voting: The Median Voter Result is an Artifact," *Journal of Economic Theory*, 16: 208-219.

Hunt, Brian R., Tim Sauer, and James A. Yorke. 1992. "Prevalence: A Translation-Invariant 'Almost Every' on Infinite-Dimensional Spaces," *Bulletin (New Series) of the American Mathematical Society* 27: 217-238.

Ledyard, J. 1984. "The Pure Theory of Large Two-Candidate Elections," *Public Choice*, 44: 7-41.

Mas-Colell, Andreu. 1985. *The Theory of General Economic Equilibrium: A Differentiable Approach*. Cambridge: Cambridge University Press.

Patty, John W. 2001. "Plurality and Probability of Victory: Some Equivalence Results," *Public Choice*, 112: 151-66.


[26] I thank Richard McKelvey for suggesting this extension.

[27] It is already known, for example, that equivalence is even more unusual in electoral models when abstention is allowed in the sense that equivalence may hold in an electoral setting without abstention but fail once abstention is allowed (e.g., Hinich (1977) and Patty (2001)).

Patty, J. W. 2000. *Voting Games with Incomplete Information*. Unpublished Ph.D. Dissertation, California Institute of Technology.

# A Proof of Lemma 1

**Lemma 1** *For any electoral game with differentiable response functions* $p$*, any candidate* $j \in J$*, and any policy profile* $x \in Y$*,*

$$D_{x_j} R_j(x) = \sum_{i \in N} \delta_i^j (p_{-i}(x)) D_{x_j} p_i^j(x).$$

*Proof*:

$$\begin{aligned} R_l(x) &= \sum_{a \in A} \frac{1}{|W(a)|} \mathbf{1}[l \in W(a)] \Pr[a|p(x)] \\ &= \sum_{a \in A : l \in W(a)} \frac{1}{|W(a)|} \prod_{i=1}^{N} p_i^{a_i}(x). \end{aligned}$$

$$\begin{aligned} D_{x_l} R_l(x) &= \sum_{a \in A : l \in W(a)} \left[ \frac{1}{|W(a)|} \sum_{i=1}^{N} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \\ &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{a \in A : l \in W(a), |W(a)|=k} \left[ \sum_{i=1}^{N} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \right] \\ &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in A : l \in W(a), |W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \right] \qquad (9) \\ &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in A : l \in W(a), |W(a)|=k, a_i = l} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{l}(x) \right. \right. \\ &\quad \left. \left. + \sum_{a \in A : l \in W(a), |W(a)|=k, a_i \neq l} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \right]. \qquad (10) \end{aligned}$$

For any voter $i \in N$ and any vector of policy proposals $x \in Y$, $\sum_{l=1}^{J} p_i^l(x) = 1$, so that, for any candidate $j \in J$, $\sum_{l=1}^{J} D_j p_i^l(x) = 0$. Rewriting Equation 10:

$$\begin{aligned} D_{x_l} R_j(x) &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in \mathcal{D}(i;l):|W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \right.\right. \\ &\quad + \sum_{a \notin \mathcal{D}(i;j):l \in W(a),|W(a)|=k,a_i=l} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \\ &\quad \left.\left. + \sum_{a \in A:l \in W(a),|W(a)|=k,a_i \neq l} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \right]. \end{aligned} \tag{11}$$

For any voter $i$, any candidate $j$, and any vote vector $a \in A$, $a_i \neq j$ implies that $a \notin \mathcal{D}(i;j)$. Thus, it is possible to combine the second and third inner sums in Equation 11 and obtain

$$\begin{aligned} D_{x_l} R_j(x) &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in \mathcal{D}(i;l):|W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \right.\right. \\ &\quad \left.\left. + \sum_{a \notin \mathcal{D}(i;j):l \in W(a),|W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^{a_i}(x) \right] \right]. \end{aligned} \tag{12}$$

For any voter $i \in N$ and candidate $j \in J$, let $\mathcal{ND}(i;j) \subset A_{-i}$ denote the set of vectors of votes other than $i$'s in which $j \in W(A)$ and $i$ can not be pivotal for $j$. That is, regardless of $i$'s vote, $W(a)$ remains the same (and includes $j$). Formally,

$$\mathcal{ND}(i;j) = \{a_{-i} \in A_{-i} : j \in W(a_i; a_{-i}) \forall a_i \in J\}.$$

Rewriting Equation 12,

$$\begin{aligned} D_{x_l} R_j(x) &= \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in \mathcal{D}(i;l):|W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \right.\right. \\ &\quad \left.\left. + \sum_{a_{-i} \in \mathcal{ND}(i;j):|W(a)|=k} \left[ \sum_{m=1}^{J} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^m(x) \right] \right] \right]. \end{aligned} \tag{13}$$

Since $\sum_{m=1}^{J} D_j p_i^m(x) = 0$ for any $i \in N$ and $x \in Y$, the second inner sum in Equation 13 vanishes, leaving

$$D_{x_l} R_j(x) = \sum_{k=1}^{J} \frac{1}{k} \left[ \sum_{i=1}^{N} \left[ \sum_{a \in \mathcal{D}(i;l);|W(a)|=k} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \right] \right].$$

Then, summing over the cardinality of $W(a)$, we obtain

$$D_{x_l} R_j(x) \ = \ \sum_{i=1}^{N} \left[ \sum_{a \in \mathcal{D}(i;l)} \frac{1}{|W(a)|} \left[ \prod_{j \neq i} p_j^{a_j}(x) \right] D_{x_l} p_i^l(x) \right]. \tag{14}$$

Finally, using Equation 1 and substituting $\delta_i^l(p_{-i}(x))$ into Equation 14, we obtain

$$D_{x_l} R_l(x) = \sum_{i=1}^{N} \delta_i^j(p_{-i}(x)) D_{x_l} p_i^l(x),$$

as was to be shown. ∎