Strategic Manipulation in Representative Institutions

Elizabeth Maggie Penn* Sean Gailmard† John W. Patty‡

November 24, 2008

Abstract

This paper considers environments in which individual preferences are single-peaked with respect to an unspecified ordering of the alternative space. We show that in these environments, any institution that is coalitionally strategy-proof must be dictatorial. Thus, any non-dictatorial institutional environment that does not explicitly utilize an *a priori* ordering over alternatives in order to render a collective decision is necessarily prone to the strategic misrepresentation of preferences by an individual or group. Accordingly, insincere behavior is inherent to the vast majority of real-world lawmaking systems, even when the policy space is unidimensional and the core is nonempty. We conclude with a discussion of the implications for claims about the political representation of interests within real-world political institutions.

*Assistant Professor, Department of Government & Institute for Quantitative Social Science, Harvard University. Email: epenn@gov.harvard.edu.

†Assistant Professor, Department of Political Science, University of California, Berkeley. Email: gailmard@berkeley.edu.

‡Assistant Professor, Department of Government & Institute for Quantitative Social Science, Harvard University. Email: jpatty@gov.harvard.edu.

1 Collective Rationality and Neutrality

The following theorem — the Gibbard-Satterthwaite theorem (henceforth, *G-S*) — is central to the study of both collective choice institutions themselves and individual behavior within such institutions.

Theorem 1. *(Gibbard, 1973), (Satterthwaite, 1975). If there are three or more feasible policy alternatives and each individual may have any preference over the alternatives, then the only collective choice institution in which no individual ever has an incentive to misrepresent his or her true preferences is a dictatorship.*

Like Arrow's Possibility Theorem, G-S depends heavily on the assumption of *unrestricted domain*, which states that all individuals may rank all alternatives however they want. In particular, both theorems leverage the fact that, in the presence of a cyclic profile of preferences, any responsive institution that produces a transitive ranking over alternatives must on occasion be called upon to break a tie in favor of a minority. Many scholars have called into question the empirical relevance of both G-S and Arrow's theorem precisely because of the unrestricted domain assumption. Mackie provides perhaps the most comprehensive case for the non-existence of majority-preference cycles in real politics, arguing that virtually every published empirical claim of a majority-preference cycle has been made in error (Mackie, 2003). Mackie claims that voter preferences in general are single-peaked over any domain of issues under consideration at a given time. The validity of this claim is not an issue we tackle in this paper. Suffice it to say that many issue spaces in politics and economics can be naturally interpreted as being one-dimensional, and it is not a stretch to assume that preferences are single-peaked over these spaces. Furthermore, it is well-known that when voter preferences *are* single-peaked with respect to a fixed ordering over alternatives, there exists a non-dictatorial institution in which every individual always has an incentive to truthfully represent his or her preferences – namely, majority rule.

Theorem 2. *(Black, 1948).*[1] *If individual preferences and ballots are single-peaked with respect to a fixed ordering, then majority rule is a non-dictatorial collective choice institution with full range in which no individual ever has an incentive to misrepresent his or her preferences.*

And so it would appear that if an issue space admits single-peaked preferences, then we have solved the question of whether preferences have been represented accurately. Or have we? What, exactly, does the existence of a well-defined majority will and nonempty core get us? Does it, for example, imply that reasonably democratic institutions will produce the core as a policy outcome? Does it imply that that individuals have an incentive to behave sincerely? As the following statements demonstrate, there is a large literature in political science that argues that single-peakedness implies a great deal:

> "When all individuals have single-peaked preference orderings the process of collective decision making is dramatically simplified." (Feld and Grofman, 1988, p. 776)

> "[When preferences are single-peaked] voters know the political world is coherently organized, the possibility of cycles is zero and the method of majority rule is wholly consistent and never tyrannical." (Riker, 1992, p. 107)

> "[R]elaxation of [unrestricted domain] provides acceptable escape-routes from Arrow's theorem and from the Gibbard-Satterthwaite theorem, compatible with all other conditions of these theorems." (Dryzek and List, 2002)

> "We do know for sure that if the distribution of preference orders is such that they are single-peaked, the Gibbard-Satterthwaite Theorem does not apply, there is no chance for strategic

[1] Black's Median Voter Theorem specifically states that in the presence of single-peaked preferences, a Condorcet winner exists. That a Condorcet winner arises as the outcome of generalized majority rule over single-peaked ballots, and that majority rule is strategy-proof in this case, was proved early on by Dummett and Farquharson (1961) and Pattanaik (1976).

voting to succeed." (Mackie, 2003, p. 161)

However, there is an important assumption implicit in our statement of Black's Theorem that is necessary in order for single-peakedness to be the "escape route" from G-S that Dryzek and List claim it to be. It requires that individuals may only vote over issues in a way that is consistent with the underlying ordering of alternatives. In other words, if the underlying ordering of alternatives on the left-right spectrum is $x < y < z$, then a strategic individual may not cast a vote for x over z and a vote for z over y. Black's Theorem then tells us that majority rule is strategy-proof provided that individuals aren't *too* strategic.[2]

In this paper we consider incentives for the strategic manipulation of institutions when individuals have preferences that are single-peaked, but are allowed to *claim* to have any preference ordering that they wish. In other words, if the true ordering of alternatives on the left-right spectrum is $x < y < z$, an individual or collection of individuals may claim to prefer z to x to y. We show that in such situations, inducing individuals to truthfully reveal their preferences is not automatic *even when preferences are single-peaked and there exists a clear majority rule core.* In particular, inducing truthful revelation at every single-peaked profile of preferences requires the use of a dictatorial collective choice procedure – a procedure that always grants decision-making authority to a unique individual. Formally, we show that any institution that is not dictatorial necessarily is not *coalitionally strategy-proof*: there exists some situation in which a person or a collection of people have an incentive to misrepresent their preferences in pursuit of a different collective choice. More importantly, this conclusion holds even in the absence of majority-preference cycles and the presence of a Condorcet winner.

[2]This point was noted by Blin and Satterthwaite (1976), who show that majority rule with Borda completion is strategy-proof when preferences and ballots are required to be single-peaked with respect to a common ordering, but is manipulable when ballots are no longer required to be single-peaked with respect to the common ordering.

1.1 Why Manipulation is Important

The question of manipulation – the misrepresentation of one's true preferences for individual gain – has been examined by social scientists for over two centuries.[3] Ignoring the unsettling mendaciousness inherent in manipulative behavior, manipulation is generally problematic for any problem relating to inference. For example, how does a legislator's roll call vote relate to his or her policy preferences? How does an individual voter's vote choice reflect his or her preferences over the parties and/or candidates? More subtly, does the composition of a committee reflect the preferences of a group's members? Does a juror's vote to convict a defendant truthfully reflect the juror's beliefs about the defendant's guilt or innocence? Does an executive's best choice of political appointees necessarily pursue the policies that the executive would individually pursue in those positions? The inference problem is also highly relevant for mechanism design issues, such as how a social planner might go about choosing the most efficient form of regulation, when the regulated firms have private information about their cost structures.

In a nutshell, we argue that the possibility of manipulation is endemic to virtually all political institutions, even in settings where it has been considered unimportant, and regardless of the attractiveness of certain policies under consideration. This is not to say that manipulation is a bad thing; in fact, some have convincingly argued that the opposite is actually true.[4] For example, Miller (1977) shows that when all individuals vote strategically, outcomes may be obtained that are Pareto superior to outcomes obtained if all individuals vote truthfully. Our argument is simply that manipulation may be problematic for those of us wishing to study the relationship between individual goals and behaviors, even in settings in which a

[3] In his 1788 treatise "On the Constitution and the Functions of Provincial Assemblies," Condorcet argued that Borda's scoring method was highly prone to strategic manipulation by voters (Young, 1995).

[4] See also Dowding and Van Hees (2007), in which the authors distinguish between "sincere" and "non-sincere" forms of manipulation.

"majority will" is well-defined.

In this respect, and before continuing on to the theory and results, it is useful to compare problems of manipulation to the "Arrovian problem" of the conditions under which one can define an unambiguous notion of "social will." This contrast is relevant because, as mentioned above, real-world institutions are subject to manipulation even if one assumes that the central difficulty with defining a minimally democratic notion of collective preference – the possibility that majority preference is cyclic – does not occur. We state and discuss Arrow's result below.

Theorem 3. *(Arrow, 1951). If there are three or more alternatives and at least two individuals, each of whom may have any preference over the alternatives, then the only Pareto efficient preference aggregation rule that satisfies independence of irrelevant alternatives is dictatorial.*

G-S does not preclude any profile of individual preferences, a direct analogue of Arrow's requirement of *unrestricted domain*. Arrow's result, which establishes that the notion of a well- (and universally-) defined "social will" is incompatible with very minimal democratic principles, is closely related to G-S and it has been demonstrated that both theorems can be obtained from what is essentially a single proof (Reny, 2000). However, the conclusions of G-S differ from those of Arrow's possibility theorem in several important ways. As opposed to the acceptance of social indifference within the Arrovian framework, G-S requires that a *unique* collective choice be generated (*i.e.*, G-S deals with "revealed collective preference"). In practical terms, this requirement is equivalent to assuming that any tie for the most-preferred choice must be broken. This requirement implies that G-S is more "institutionally constructive" than Arrow: while Arrow's result maps preference profiles into social preferences, G-S requires that the social preference generate a non-manipulable choice. In other words, G-S not only requires that ties be broken – it requires that they be broken in a way that does not strictly reward insincere behavior by the voters.

There is another important difference between Arrow's theorem and G-S that is particularly relevant when considering preference domain restrictions. In Arrow's theorem, institutions consider only sincere profiles of preferences; in the G-S theorem, they do not. Thus, a preference domain restriction (e.g. single-peakedness) is a stronger restriction in Arrow's theorem than under the G-S theorem, unless an institution under consideration specifically requires *ballots* to be drawn from the same domain as *preferences*. The differential effect of preference domain restrictions in the two theorems plays out in our results. Our proof that coalitional non-manipulability requires dictatorship on single-peaked preference domains leans heavily on a related, but normatively weaker, result that we have proved for Arrow's theorem (Gailmard et al., 2008): Any weakly Paretian preference aggregation rule that is independent of irrelevant alternatives must be *neutral*, even when preferences are known to be single-peaked. In other words, even in instances in which there is a well-defined, transitive majority preference relation, neutrality is required for collective choice to be simultaneously transitive, weakly Paretian and independent of irrelevant alternatives. In our extension of this result to G-S we can replace neutrality with dictatorship precisely because of the fact that the single-peaked preference domain restriction is weakened by assuming that institutions are required to take *all* ballot profiles as an input.

The key to both results is that knowing with certainty that the alternatives *can* be ordered so as to induce a single-peaked profile of preferences is not equivalent to knowing *how* the alternatives will be ordered. Thus, single-peakedness in and of itself is not "enough" information to enable a non-neutral institution to produce an IIA and weakly Paretian social ranking over alternatives, and it is not enough information to enable a non-dictatorial institution to be coalitionally strategy-proof. Given the generality and power of Arrow's Theorem and the Gibbard-Satterthwaite Theorem, the novelty of our results lies in their application to a canonical setting for models of political institutions: the unidimensional spatial model.[5] Furthermore, the unidimensional spatial model is often invoked precisely because it is considered to be immune from the conclusions of Arrow and G-S. In this paper we argue that these conclusions can be interpreted as institutional problems, rather than problems stemming from underlying majority preference cycles. In other words, violating conditions such as Pareto efficiency, independence of irrelevant alternatives, transitivity of collective choice, and strategy-proofness is an irresolvable consequence of the vast majority of collective decision-making procedures in use in the world, and poses problems that are separate from the underlying structure of preferences.

The following section defines the theoretical framework that we utilize in both results. Section 3 proves our main results: that coalitional strategy-proofness on the domain of single-peaked preferences first implies neutrality, and then dictatorship. Section 4 presents several examples of the applicability of our results to real-world institutions, including binary voting procedures and institutions with a status quo bias. This section also briefly discusses several normative implications of our results, including questions of optimal delegation and the desirability of Pareto efficiency in certain decision-making procedures. In Section 5, we conclude and offer a brief discussion of the connections between our results and the analysis of representative institutions in general.

2 Notation and Definitions

There is a finite collection of K *alternatives* (or *policies*), X, and a finite collection of n *individuals* (or *voters*) N. We assume that $K \geq 3$ and $n \geq 2$. Individual i's preferences are represented by a strict,

[5] A smattering of examples to justify the term "canonical" might include Downs (1957), Davis et al. (1970), McCubbins et al. (1994), Poole and Rosenthal (1997), and Krehbiel (1998) among, of course, many others.

transitive and complete binary relation P_i. The notation xP_iy implies that i strictly prefers x to y. If $x_i^*P_iy$ for all $y \neq x_i^*$, then x_i^* is referred to as i's *most-preferred policy* or *ideal point*. Let $x_i^*(\rho)$ be i's ideal point under profile ρ.

Throughout, $\rho = (P_1, ..., P_n)$ denotes an n-dimensional preference profile describing the preferences of all individuals: the notation $\mathcal{P}^n$ represents the collection of all n-dimensional profiles of strict orders on X. Any nonempty set $\mathcal{D} \subseteq \mathcal{P}^n$ is referred to as a *preference domain*, and with strict inclusion, $\mathcal{D}$ is referred to as a *restricted domain*. We will come back to restricted domains in more detail in Section 2.3. For any preference profile $\rho \in \mathcal{P}^n$, $\rho|_S$ denotes the restriction of ρ to the set of alternatives $S \subseteq X$. Similarly, for any individual preference $P \in \mathcal{P}$, $P_i|_S$ denotes the restriction of i's preference relation to the set S. For any preference profile $\rho \in \mathcal{P}^n$ and pair of alternatives $(x, y) \in X^2$, the notation $P(x, y; \rho) \equiv \{i \in N : xP_iy\}$ denotes the set of individuals who strictly prefer x to y under ρ.

2.1 Preference Aggregation Rules

Letting $\mathcal{R}$ be the collection of weak orders over X, a *preference aggregation rule* is any function, $F : \mathcal{D} \to \mathcal{R}$, that maps a strict preference profile from domain $\mathcal{D}$ into a weak order over X. The notation $xR_F(\rho)y$ denotes weak social preference under F at profile $\rho \in \mathcal{P}^n$ and $xP_F(\rho)y$ denotes strict social preference. The following definitions characterize several properties of preference aggregation rules.

Definition 1 (Weakly Paretian)**.** *A preference aggregation rule F is* weakly Paretian *if for all $\rho \in \mathcal{D}$ and all $(x, y) \in X^2$,*

$$P(x, y; \rho) = N \Rightarrow xP_F(\rho)y.$$

Definition 2 (Neutral)**.** *A preference aggregation rule F is* neutral *if for every $x, y, a, b \in X$, and every*

$\rho, \rho' \in \mathcal{D}$,

$$P(x, y; \rho) = P(a, b; \rho') \textit{ implies } xR_F(\rho)y \Leftrightarrow aR_F(\rho')b.$$

Definition 3 (Independent of Irrelevant Alternatives (IIA))**.** *A preference aggregation rule F is* independent of irrelevant alternatives *(IIA) if, for all $(x, y) \in X^2$ and all $\rho, \rho' \in \mathcal{D}$,*

$$\rho|_{\{x,y\}} = \rho'|_{\{x,y\}} \Rightarrow F(\rho)|_{\{x,y\}} = F(\rho')|_{\{x,y\}}.$$

2.2 Collective Choice Functions

A *collective choice function*, or *choice function*, is any function, $\phi : \mathcal{P}^n \rightarrow X$ that maps any strict profile of orderings over alternatives into X. Throughout, we will assume that ϕ has *full range*: for any $x \in X$, there exists a $\rho \in \mathcal{P}^n$ such that $\phi(\rho) = x$. Thus, a preference aggregation rule produces an ordering over the elements of X while a choice function simply produces a single outcome. The notation $\phi(\rho) = x$ says that choice function ϕ produces outcome x at profile ρ.

While we require choice functions to map any strict profile into a social outcome, we do not require individuals' true preference orderings to be drawn from the full set $\mathcal{P}^n$. This is because we are interested in the, possibly insincere, behavior induced by a choice function when true individual preferences are drawn from a restricted domain. We call the *preference domain* of a choice function $\mathcal{D}$, while the *ballot domain* of all choice functions is assumed to be $\mathcal{P}^n$.[6] In other words, while true individual preferences may come from a restricted set of orderings, individual behavior is only required to be individually rational, in the sense of being rationalizable by a transitive binary relation. Throughout, we will use the notation (P'_i, ρ_{-i}) to denote a ballot profile in which i submits ballot P'_i, and all others submit ballots as in profile ρ. More generally, the

[6] In Section 4.2 we discuss the implications of restricting the ballot domain for a specific class of institutions with an agenda setter.

notation (ρ'_L, ρ_{-L}) denotes a ballot profile in which all members $i \in L \subseteq N$ submit ballots as under ρ', and all individuals not in L submit ballots as under ρ.

The following definitions characterize several properties of collective choice functions.

Definition 1′ (Weakly Paretian)**.** *A collective choice function ϕ is* weakly Paretian *if for all $\rho \in \mathcal{D}$ and all $(x, y) \in X^2$,*

$$P(x, y; \rho) = N \Rightarrow \phi(\rho) \neq y.$$

Definition 2′ (Neutral)**.** *A collective choice function ϕ is* neutral *if for every permutation $\sigma : X \to X$, and every profile $\rho \in \mathcal{D}$,*

$$\phi(\rho) = x \Leftrightarrow \phi(\sigma(\rho)) = \sigma(x).$$

Definition 4 (Monotonic)**.** *A collective choice function ϕ is* monotonic *if, for all $(x, y) \in X^2$ and all $\rho, \rho' \in \mathcal{D}$,*

$$P(x, y; \rho) \subseteq P(x, y; \rho') \text{ and } \phi(\rho) = x \Rightarrow \phi(\rho') \neq y.$$

Definition 5 (Dictatorial)**.** *A collective choice function ϕ is* dictatorial *if for some $i \in N$ and for all $\rho \in \mathcal{D}$,*

$$\phi(\rho) = x_i^*(\rho),$$

where $x_i^*(\rho)$ is i's ideal point under profile ρ.

Definition 6 (Strategy-Proof (SP))**.** *A collective choice function ϕ is* manipulable *if, for some $\rho = (P_1, ..., P_n) \in \mathcal{D}$ and $i \in N$ there exists a $P_i' \in \mathcal{P}$ such that*

$$\phi(P_i', \rho_{-i}) P_i \phi(\rho).$$

A choice function is strategy-proof *if it is not manipulable.*

Definition 7 (Coalitionally Strategy-Proof (CSP))**.** *A collective choice function ϕ is* coalitionally manipulable *if, for some $\rho = (P_1, ..., P_n) \in \mathcal{D}$ and $L \subseteq N$ there exists a $\rho' \in \mathcal{P}^n$ such that*

$$\phi(\rho'_L, \rho_{-L}) P_i \phi(\rho) \text{ for all } i \in L.$$

A choice function is coalitionally strategy-proof *if it is not coalitionally manipulable.*

Note that individually manipulable social choice functions are coalitionally manipulable (by a coalition of one). However, the converse need not be true; there may be instances in which a social choice function may only be manipulable by a sufficiently large coalition. When $\mathcal{D} = \mathcal{P}^n$, then individual non-manipulability implies coalitional non-manipulability, because it implies dictatorship. Thus, in a setting with unrestricted domain, individual and coalitional manipulability and non-manipulability are equivalent.

2.3 Single-Peaked Preferences

In this section we define the domain of single-peaked preferences. This domain has attracted the interest of many scholars because it has been shown to lead to the existence of non-dictatorial Arrovian preference aggregation rules and, when ballots are required to be single-peaked, to non-manipulable, non-dictatorial collective choice functions. Our interest is less about the existence, and more about the characterization, of such preference aggregation rules and choice functions on this restricted domain.

Single-Peaked Preferences. The domain of *single-peaked preferences* is the set of all profiles of preferences such that there exists a function $Q : X \rightarrow \{1, 2, \ldots, K\}$ such that Q is a bijection and every individual's preferences are consistent with a quasi-concave utility function of $\{Q(x) : x \in X\}$. We denote the single-peaked preference domain by $\mathcal{S}^n \subset \mathcal{P}^n$. We will at times refer to the ordering that profile $\rho \in \mathcal{S}^n$

is single-peaked with respect to as Q_ρ. When referring to this ordering, if alternative x is above y with respect to Q we write $x >_Q y$.

While this preference restriction is widely utilized and intuitively quite simple, Ballester and Haeringer (2007) prove that the set $\mathcal{S}^n$ is completely characterized by two conditions, worst-restriction[7] and α-restriction, both defined below.

Definition 8 (Worst-restriction)**.** *A profile ρ is* worst-restricted *if, for every triple of alternatives, $(x, y, z) \in X^3$, there exists an $a \in \{x, y, z\}$ such that for all $i \in N$, $a \succ_i b$ for some $b \in \{x, y, z\} \setminus \{a\}$.*

In words, a profile is worst-restricted if for every triple $(x, y, z) \in X^3$, there is some element of that triple that no individual ranks last relative to the other two elements of the triple.

Definition 9 (α-Restriction)**.** *A preference profile ρ is α-*restricted *if there do not exist two agents, $i, j \in N$, and four alternatives w, x, y, and z such that*

1. *The preferences over w, x, and z are opposite: wP_ixP_iz and zP_jxP_jw.*
2. *The players agree about the ranking of y and x: yP_ix and yP_jx.*

Definition 10 (Single-Peakedness)**.** *A preference profile is* single-peaked *if and only if it satisfies worst-restriction and $\alpha-$restriction (Ballester and Haeringer, 2007).*

It is important to note at this point that the domain $\mathcal{S}^n$ is the set of *all* single peaked preference profiles. In other words, in *a priori* terms, any ordering of the alternatives is possible.[8]

[7]See Sen (1966) and Sen and Pattanaik (1969) for a more thorough discussion of worst-restriction.

[8]This point is a technical one, but important for broader considerations of the results in this paper. In particular, for any given linear ordering of the alternatives, $Q \in \mathcal{P}$, one can identify the set of preferences that are single-peaked with respect to $\mathcal{Q}$, this set is denoted by $\mathcal{S}_Q$, and the set of all profiles of such preferences is denoted by $\mathcal{S}_Q^n$. This space is widely discussed in the political

Ubeda (2003) has recently used a different domain restriction, the 2-free triple domain ($\mathcal{T}_2^n$),[9] to demonstrate that on any domain satisfying this restriction, weak Pareto and IIA imply neutrality, a conclusion that mirrors our own (Theorem 4, below). The key distinction between Ubeda's result and Theorem 4 is that the 2-free triple domain and the single-peaked domain are not nested. Specifically, for all $n \geq 2$, $\mathcal{T}_2^n \not\subseteq \mathcal{S}^n$ and $\mathcal{S}^n \not\subseteq \mathcal{T}_2^n$. In other words, satisfaction of either the 2-free triple restriction or single-peakedness does not imply satisfaction of the other. With these preliminaries in hand, we are now in a position to state and prove our main results.

3 Stability and Coalitional Strategy-Proofness on Single-Peaked Domains

To prove that coalitional strategy-proofness implies neutrality and then dictatorship on domain $\mathcal{D} = \mathcal{S}^n$, we utilize the following lemmas and theorem.

Lemma 1. *Let ϕ be a coalitionally strategy-proof collective choice function. If $\mathcal{D} = \mathcal{S}^n$, then ϕ is weakly Paretian.*

Proof: Consider a $\rho \in \mathcal{S}^n$ such that for all $i \in N$, xP_iy, but $\phi(\rho) = y$. By full range, $\exists \rho' \in \mathcal{P}^n$ with $\phi(\rho') = x$. Then $\phi(\rho')P_i\phi(\rho)$ for all $i \in N$, and ϕ is manipulable by coalition N. It follows that $\phi(\rho) \neq y$ if ϕ CSP. $\square$

economy literature. For a succinct and lucid overview of the power of the assumption that Q is known *a priori*, see Chapter 2.4 of Austen-Smith and Banks (2004), in particular Theorem 2.4.

[9] This domain restriction says that for any triple of alternatives, only two orderings of the triple are possible across all individuals. While profiles on this domain will satisfy worst-restriction, they may fail α-restriction, with a clear example being the case with two individuals with preferences: $wP_1yP_1xP_1z$ and $zP_2yP_2xP_2w$. Similarly, the following 3-player profile is single-peaked but is not an element of the 2-free triple domain: xP_1yP_1z, yP_2xP_2z, and zP_3xP_3y.

Lemma 2. *Let ϕ be a coalitionally strategy-proof collective choice function. If $\mathcal{D} = \mathcal{S}^n$, then ϕ is monotonic.*

Proof: Suppose that ρ and ρ' are single-peaked, but violate monotonicity, with $\phi(\rho) = x$, $\phi(\rho') = y$, and $P(x, y; \rho) \subseteq P(x, y; \rho')$. We will show that this implies ϕ is not CSP on $\mathcal{S}^n$. Throughout the proof, let $P(x, y; \rho) = A$ and $P(y, x; \rho') = B$. We know that $A \cap B = \emptyset$.

First, to simplify notation, change ρ so that for all i such that $x\overline{P}_i y$ under ρ, every $\overline{P}_i$ is replaced by an identical new ordering P_i with x top-ranked and y as high in i's ranking as possible while maintaining P_i single-peaked with respect to Q_ρ. It is easy to verify that such an ordering exists. Similarly, change ρ' so that for all j with $y\overline{P}'_j x$ under ρ', y is top-ranked with respect to the new P'_j, and x is as high as possible while maintaining that P'_j be single peaked with respect to $Q_{\rho'}$. CSP implies that for each of these new profiles (which in an abuse of notation we will still call ρ and ρ') $\phi(\rho) = x$ and $\phi(\rho') = y$.

Consider the triple, $x, y, z \in X$, with z arbitrary. $\rho, \rho' \in \mathcal{S}^n$ imply that for each profile, one of only two elements of this triple may be lowest-ranked by any individual. If at profile ρ these two elements are a, b, we say that (a, b) is lowest-ranked for $\rho|_{\{a,b,z\}}$. Note that in constructing ρ above, we have ensured that the only instances in which zP_iy for $i \in A$ are those in which z lies between x and y under the ordering Q_ρ (i.e. $x >_{Q_\rho} z >_{Q_\rho}> y$, or the reverse). Similarly, zP'_jx for $j \in B$ implies that z lies between x and y according to $Q_{\rho'}$.

Construct a new profile $\rho^* \in \mathcal{P}^n$ in which $P^*_i = P_i$ for $i \in A$, and $P^*_j = P'_j$ for $j \in B$. Note that rankings are unspecified for $k \notin A \cup B$, if such individuals exist. These individuals' transitive rankings can be arbitrarily assigned. Also note that ρ^* may not be single-peaked; our choice function is still required to produce an outcome at this profile, however. Then it must be the case that $\phi(\rho^*) \notin \{x, y\}$, else either coalition $N \setminus A$ could manipulate ρ with ρ^*, or $N \setminus B$ could manipulate ρ' with ρ^*. Thus, $\phi(\rho^*) = z$.

Case 1: (x, y) is lowest-ranked for either $\rho|_{\{x,y,z\}}$ or for $\rho'|_{\{x,y,z\}}$. Without loss of generality, assume that (x, y) is lowest-ranked for $\rho|_{\{x,y,z\}}$. This implies that for all individuals $k \notin A$, $yP_kx \Rightarrow zP_kx$. Thus, ϕ is manipulable at ρ by coalition $N \setminus A$ submitting ballots as in ρ^*; these individuals can guarantee themselves the outcome z, which they all prefer to x. It follows that ϕ is not CSP.

Case 2: z is a lowest ranked element of both $\rho|_{\{x,y,z\}}$ and $\rho'|_{\{x,y,z\}}$. Note that by the construction of ρ and ρ' above, yP_iz for all $i \in A$ and xP'_jz for all $j \in B$.[10]

To recap, we have that all members of A (resp. B) have the same preference ordering over all alternatives, and that this ordering ranks xP_iyP_iz (resp. yP_jxP_jz). Note also that x, y, z need not appear consecutively in these individuals' rankings. Furthermore, ρ^* as constructed above still yields z as its choice. However, now members of coalition $N \setminus A$ (resp. $N \setminus B$) may rank z last relative to x and y, and may not have an incentive to manipulate ϕ by submitting ballots as in ρ^*.

We will now construct a new $\rho^o \in \mathcal{S}^n$ and show that ϕ CSP for ρ implies that ϕ not CSP for ρ^o. This will take several intermediate steps. First, consider the ordering over alternatives induced by the preferences of individuals in A under ρ; we will call this ordering Q_A, and we know that under this ordering $x >_{Q_A} y >_{Q_A} z$. Construct a new profile $\hat{\rho}$ where $\hat{P}_i = P_i$ for all $i \in A$. Clearly this $\hat{P}_i$ is single-peaked with respect to the ordering Q_A, as it *is* this ordering. For all $j \notin A$, assign each member of j an identical preference ordering $\hat{P}_j$ that ranks y at the top of the ballot, ranks $z\hat{P}_jx$, and is single-peaked with respect to Q_A. Such an ordering exists because y lies between z and x according to Q_A. Thus, $\hat{\rho} \in \mathcal{S}^n$.

[10]This immediately implies that ϕ does not satisfy Pareto efficiency on the full domain $\mathcal{P}^n$, because ρ^* can be constructed so as to have all individuals not in $A \cup B$ place x and y at the top of their ballots, and so $x, y > z$ on all ballots ρ^*_i and yet $\phi(\rho^*) = z$.

We know that coalition $N \setminus A$ can submit ballots as in ρ^* and receive z as an outcome, which they prefer to x. Thus, $\phi(\hat{\rho}) \neq x$. Furthermore, $\phi(\hat{\rho}) \neq z$, by weak Pareto, because for all $i \in N$, $y\hat{P}_i z$. And by CSP of ρ, $\phi(\hat{\rho}) \neq y$, because then ρ would be manipulable by coalition $N \setminus A$ submitting ballots as in $\hat{\rho}$. It follows that $\phi(\hat{\rho}) = w$, with all members $j \in N \setminus A$ ranking $y\hat{P}_j w\hat{P}_j z\hat{P}_j x$. However, this implies that for all $i \in A$, $w\hat{P}_i y$, otherwise $\hat{\rho}$ would be manipulable by coalition A submitting ballots identical to those for $N \setminus A$, and ensuring an outcome of y by weak Pareto. Thus Q_A ranks the alternatives $x >_{Q_A} w >_{Q_A} y >_{Q_A} z$.

Now consider a new profile $\hat{\rho}^1$ in which everyone in A ranks the alternatives as in $\hat{\rho}$. For all $j \in N \setminus A$, assign each j an identical preference ordering that ranks y at the top, ranks $z\hat{P}^1_j w\hat{P}^1_j x$, and is single-peaked with respect to Q_A. Again, such an ordering exists given that Q_A ranks the alternatives $x >_{Q_A} w >_{Q_A} y >_{Q_A} z$. Thus, $\hat{\rho}^1 \in \mathcal{S}^n$. Using the same logic as above, we get that $\phi(\hat{\rho}^1) \neq x$ or w (else $N \setminus A$ manipulates with ρ^* to get z), that $\phi(\hat{\rho}^1) \neq z$ by Pareto efficiency, and that $\phi(\hat{\rho}^1) \neq y$ by CSP of our original ρ. Thus, $\phi(\hat{\rho}^1) = w^1$, with individuals $j \in N \setminus A$ ranking $y\hat{P}^1_j w^1\hat{P}^1_j z\hat{P}^1_j w\hat{P}^1_j x$. Again, it follows that for all $i \in A$, $\hat{P}^1_i$ ranks $w^1\hat{P}^1_i y$, otherwise this coalition would manipulate $\hat{\rho}^1$ in order to get y as the outcome. Thus, Q_A ranks the alternatives $x >_{Q_A} \{w^1, w\} >_{Q_A} y >_{Q_A} z$ (the w^1, w ranking need not be specified in the proof, although by coalition $N \setminus A$'s preferences, the ranking must have $w >_{Q_A} w^1$).

Repeat the above steps for $k = 2, ..., |X| - 4$ by choosing a new $\hat{\rho}^k$ with $y\hat{P}^k_j z\hat{P}^k_j...$ for all $j \in N \setminus A$, and $\hat{P}^k_i = P_i$ for $i \in A$. Such a $\hat{\rho}^k$ can be constructed so as to always remain single-peaked with respect to Q_A because it must always be the case that the social choice at each stage, w^k, is ranked between x and y for all $i \in A$. We ultimately get that Q_A ranks the alternatives $x >_{Q_A} \{w, w^1, w^2, ...w^{K-4}\} >_{Q_A} y >_{Q_A} z$. At this point, construct ρ^o so that P^o_j is such that $yP^o_j zP^o_j \{w, w^1, ..., w^{|X|-4}\}P^o_j x$ for all $j \in N \setminus A$ and $P^o_i = P_i$. Again, ρ^o is single-peaked with respect to Q_A. CSP requires $\phi(\rho^o)P^o_j z$ for all $j \in N \setminus A$, otherwise this coalition would manipulate with ballots as in ρ^*. CSP of ϕ at ρ requires $\phi(\rho^o) \neq y$. And these two statements imply a contradiction. Thus, ϕ is not CSP. It follows that ϕ CSP implies ϕ monotonic when $\mathcal{D} = \mathcal{S}^n$. $\square$

Theorem 4. *Let F be a weakly Paretian and IIA preference aggregation rule. If $\mathcal{D} = \mathcal{S}^n$, then F is neutral.*

Proof: See Gailmard et al. (2008).

Theorem 5. *Let ϕ be a coalitionally strategy-proof collective choice function. If $\mathcal{D} = \mathcal{S}^n$, then ϕ is neutral.*

Proof: We will prove the result by showing that we can use ϕ to construct a preference aggregation rule $F : \mathcal{S}^n \rightarrow \mathcal{R}$ that satisfies IIA and Pareto efficiency. By Theorem 4 this implies F is neutral on $\mathcal{S}^n$, which will imply ϕ is also neutral on $\mathcal{S}^n$.

First, let Q be a fixed ordering over the alternatives in X. For any $\rho \in \mathcal{S}^n$ and any $Y \subseteq X$, let ρ^Y be a new profile constructed by moving the alternatives in Y to the top of every individual's preference ordering, maintaining the individual orderings on Y as specified by ρ, and imposing the ordering given by Q on all other alternatives in every individual's profile. Thus, for all $i \in N$ and all $x, y \in Y$, $xP_iy \Rightarrow xP_i^Y y$ and $yP_ix \Rightarrow yP_i^Y x$, and $x, yP_i^Y z$ for all $i \in N$, $z \notin Y$. ρ^Y is clearly single-peaked, as its construction does not violate α-restriction or worst-restriction. $\phi(\rho^Y)$ chooses an element of Y by Pareto efficiency.

For any $\rho \in \mathcal{S}^n$ define a complete binary relation $R(\rho)$ such that $xR(\rho)y \Leftrightarrow x \in \phi(\rho^{\{xy\}})$. R is IIA on the domain $\mathcal{S}^n$ by its definition: for any two profiles ρ, ρ' with the same pairwise rankings over x, y, $\rho^{\{xy\}} = \rho'^{\{xy\}}$. R is also Pareto efficient on $\mathcal{S}^n$, by the Pareto efficiency of ϕ.

Suppose $\rho \in \mathcal{S}^n$ and $xR(\rho)yR(\rho)z$. We must show that $xR(\rho)z$ (i.e. that R is transitive on the domain $\mathcal{S}^n$). First, note that $\phi(\rho^{\{xyz\}}) = x$. If not, then by monotonicity $\phi(\rho^{\{xyz\}}) = y \Rightarrow \phi(\rho^{\{xy\}}) = y$, a contradiction. Similarly by monotonicity, $\phi(\rho^{\{xyz\}}) = z$ implies that $\phi(\rho^{\{yz\}}) = z$, another contradiction. Thus, $\phi(\rho^{\{xyz\}}) = x$, which by monotonicity implies $\phi(\rho^{\{xz\}}) = x$, and thus, that $xR(\rho)z$. Therefore $R(\rho)$

is transitive for all $\rho \in \mathcal{S}^n$. R then equals R_F for a preference aggregation rule, F with domain $\mathcal{S}^n$.

We now have a preference aggregation rule F over a set X (with $|X| \geq 3$) that is transitive, Pareto-efficient, and satisfies IIA. It follows that F is neutral on $\mathcal{S}^n$, by Theorem 4.

Our final step is to show that F neutral on $\mathcal{S}^n$ implies ϕ neutral on $\mathcal{S}^n$. Consider two profiles ρ_1 and ρ_2 where ρ_1 and ρ_2 are identical up to a relabeling of x and y.[11] To show ϕ is neutral it suffices to show that $\phi(\rho_1) = x \Rightarrow \phi(\rho_2) = y$, $\phi(\rho_1) = y \Rightarrow \phi(\rho_2) = x$, and $\phi(\rho_1) = a \Rightarrow \phi(\rho_2) = a$, for $a \neq x, y$.

First, by monotonicity, $\phi(\rho_1) = x$ implies that $xR_F(\rho_1)a$, for all $a \neq x$. By neutrality of F, we also know that $yR_F(\rho_2)a$ for all $a \neq y$, as x and y are identical under ρ_1 and ρ_2 up to a relabeling of their names. Thus, $y = \phi(\rho_2^{\{ay\}})$ for all $a \neq y$. By monotonicity, this implies that $y = \phi(\rho_2)$.

The same argument proves that $\phi(\rho_1) = a$ implies that $\phi(\rho_2) = a$, and that $\phi(\rho_1) = y$ implies that $\phi(\rho_2) = x$. Thus, ϕ CSP on $\mathcal{S}^n$ implies ϕ neutral. $\square$

For the next lemma we will need the following two definitions:

Definition 11 (Blocking coalition for (x, y))**.** *A coalition $L \subseteq N$ is a blocking coalition for (x, y) if for all $\rho = (P_1, ..., P_n) \in \mathcal{D}$ such that yP_ix for all $i \in L$ and xP_jy for all $j \notin L$, $\phi(\rho) \neq x$.*

Definition 12 (Blocking coalition)**.** *A coalition $L \subseteq N$ is a blocking coalition if for all $\rho = (P_1, ..., P_n) \in \mathcal{D}$ and all pairs $(a, b) \in X^2$, bP_ia for all $i \in L \Rightarrow \phi(\rho) \neq a$.*

Lemma 3. *ϕ coalitionally strategy-proof implies that if there exists one $\rho \in \mathcal{S}^n$ with yP_ix for all $i \in L$ and xP_jy for all $j \notin L$ and $\phi(\rho) = y$, then L is a blocking coalition.*

[11] As in the proof of our previous theorem, we can get from any profile ρ to any permutation of ρ through a series of pairwise switches.

Proof: Let ρ be such that yP_ix for all $i \in L$ and xP_jy for all $j \notin L$ and $\phi(\rho) = y$. We will first show that L is blocking for (x, y), and then that L is a blocking coalition.

Suppose that L is not blocking for (x, y), so that there exists a $\rho' \in \mathcal{S}^n$ with yP'_ix for all $i \in L$ and xP'_jy for all $j \notin L$ and $\phi(\rho') = x$. By monotonicity, this implies that $\phi(\rho) \neq y$, a contradiction. Thus, L is blocking for (x, y).

L blocking for (x, y) implies that L is blocking for all pairs (a, b), by neutrality. By monotonicity, we will show that L blocking for (a, b) implies that at any profile $\rho \in \mathcal{S}^n$ in which bP_ia for all $i \in L$, then $\phi(\rho) \neq a$. Suppose not; assume that $\phi(\rho) = a$. Let Q be the ordering that ρ is single-peaked with respect to.

Now consider a $\rho' \in \mathcal{S}^n_Q$ where for each $j \notin L$, P_j is replaced by P'_j, in which a is top-ranked under P'_j. By monotonicity, $\phi(\rho') = a$. However, under ρ' we have bP_ia for all $i \in L$ and aP_jb for all $j \notin L$. $\phi(\rho') = a$ contradicts L blocking for (a, b). Thus, L is a blocking coalition. $\square$

Theorem 6. *Let ϕ be a coalitionally strategy-proof collective choice function. If $\mathcal{D} = \mathcal{S}^n$, then ϕ is dictatorial.*

Proof: Consider three profiles $\rho_1, \rho_2, \rho_3 \in \mathcal{S}^n$ in which the alternatives x, y, z are at the top of each person's preference ordering, and all other alternatives are ordered according to a fixed ordering $Q_{-\{xyz\}}$. Thus, save for alternatives $\{x, y, z\}$, rankings over all other alternatives are identical across all individuals and all three profiles. By monotonicity, we can consider such profiles without loss of generality.

Let $L \in \mathcal{L}$ be a "minimal" blocking coalition. Thus, for any $i \in L$, the set $L \setminus \{i\}$ is not a blocking coalition. Such a coalition exists because, by Pareto, we know that the collection of blocking coalitions is nonempty. Define $\{x, y, z\}$ rankings under ρ_1, ρ_2, ρ_3 as follows:

	ρ_1	ρ_2	ρ_3
i	$x \succ z \succ y$	$x \succ y \succ z$	$x \succ y \succ z$
$L \setminus \{i\}$	$y \succ z \succ x$	$y \succ x \succ z$	$y \succ z \succ x$
$N \setminus L$	$z \succ x \succ y$	$z \succ x \succ y$	$z \succ y \succ x$

We know the following: $\phi(\rho_2) \neq z$ because all in L prefer y to z; $\phi(\rho_3) \neq z$ because all in L prefer y to z; $\phi(\rho_1) \neq y$ because everyone *not* in $L \setminus \{i\}$ prefers z to y, and we have assumed that $L \setminus \{i\}$ is not a blocking coalition; $\phi(\rho_2) \neq y$ because everyone *not* in $L \setminus \{i\}$ prefers x to y, and we have assumed that $L \setminus \{i\}$ is not a blocking coalition.

Condensing the above paragraph, we now know that $\phi(\rho_1) = x$ or z, that $\phi(\rho_2) = x$, and that $\phi(\rho_3) = x$ or y.

Case 1: First, suppose that $\phi(\rho_1) = z$. This implies that $\phi(\rho_3) = y$, because (x, z) preferences are identical across ρ_1 and ρ_3. Thus, $\phi(\rho_3) = x$ would violate monotonicity.

Now consider an insincere ballot $\hat{\rho}$ that is identical to ρ_1, ρ_2, ρ_3 for all $w \notin \{x, y, z\}$ (i.e. these alternatives are, for every individual, ordered according to $Q_{-\{xyz\}}$), and with a Condorcet cycle over x, y, z at the top:

	$\hat{\rho}$
i	$x \succ y \succ z$
$L \setminus \{i\}$	$y \succ z \succ x$
$N \setminus L$	$z \succ x \succ y$

We know that $\phi(\hat{\rho}) = d \notin \{x, y, z\}$, otherwise ρ_1, ρ_2 or ρ_3 would be manipulable by an individual or coalition submitting ballots as in $\hat{\rho}$: $\phi(\hat{\rho}) = x \Rightarrow i$ manipulates ρ_1, $\phi(\hat{\rho}) = y \Rightarrow L \setminus \{i\}$ manipulates ρ_2, and $\phi(\hat{\rho}) = z \Rightarrow N \setminus L$ manipulates ρ_3.

Now, construct a new profile $\hat{\rho}_1 \in \mathcal{S}^n$ with the preferences of all $j \neq i$ identical to those given by $\hat{\rho}$. Player i's new preferences rank $d \succ x \succ z \succ y \succ$ ..., with i's rankings over all $w \notin \{d, x, y, z\}$ unchanged. This profile is single-peaked according to the ordering specified by i's preferences, which can be verified by considering $\alpha-$ and worst-restriction. Since all players have identical orderings over alternatives not in $\{d, x, y, z\}$, moving d to the top of i's ranking cannot break worst-restriction (as all other players have the same ranking of d and any other alternative $a \neq d$), and cannot break $\alpha-$restriction, as only player i has preferences over a triple that are the reverse of another player's preferences, and i is the unique player with d at the top of his ballot.

$\phi(\hat{\rho}_1) = d$, otherwise $\phi(\hat{\rho}_1)$ would be manipulable by i submitting a ballot as in $\hat{\rho}$. But, by Lemma 3, this implies that i is a blocking coalition, and thus, a dictator, because all other players prefer x to d.

Case 2: Now suppose that $\phi(\rho_1) = x$. Then immediately, by Lemma 3, this implies that i is a dictator, because i is the unique person who prefers x to z at profile ρ_1. $\square$

4 Examples of Manipulation in One Dimension

In this section, we briefly consider three well-known models of policymaking in unidimensional spaces, and demonstrate how, in each instance, the incentive to manipulate manifests itself in the predicted collective choice of the legislature. Our examples are intended to capture several canonical settings in political science: voting over an amendment agenda, the Romer-Rosenthal *setter model* (Romer and Rosenthal, 1978), and the *ally principle* (Bendor and Meirowitz, 2004) in models of delegation.

These examples demonstrate several features of our results that may not be apparent from the theoretical sections of this paper, and that are important to note. First, as discussed in the introduction, domain restrictions pose different challenges when considering Arrow's theorem versus G-S. While coalitional strategy-proofness implies dictatorship of choice functions on $\mathcal{S}^n$, IIA and weak Pareto only imply neutrality of preference aggregation rules on $\mathcal{S}^n$. Section 4.1 illustrates this distinction by providing an example of a generally non-neutral institution that happens to be neutral on $\mathcal{S}^n$, that satisfies IIA and weak Pareto on $\mathcal{S}^n$, and that is also manipulable on $\mathcal{S}^n$.

Second, our theoretical framework leans on an assumption that individuals may submit ballots that are not single-peaked with respect to the true underlying ordering of alternatives. In Section 4.2 we depart from our current framework to provide an example of a well-known institution that is manipulable even when ballots are required to be single-peaked with respect to the same underlying ordering as preferences. It is well-known that in such settings, every strategy-proof collective choice mechanism can be represented by an augmented median voter rule, and that such mechanisms may or may not be neutral.[12] In this section we explain our example with a short proof demonstrating that augmented median voter rules are incapable of capturing many institutional settings with a status quo bias. In Section 4.3 we provide a different interpretation of manipulation that assumes that political actors (voters, legislators) may delegate policymaking authority to others. We show that, even when the policy space is one-dimensional, there are instances in which individuals or groups will knowingly grant decision-making authority to people with preferences that are different than their own, and that this can be interpreted as a form of ballot manipulation.

We do not intend for any of these examples to be surprising; in fact, the results will appear quite obvious to any reader familiar with the one-dimensional spatial model. Our intention is rather to demonstrate that

[12] Austen-Smith and Banks (2004, p. 37) provide an excellent treatment of augmented median voter rules. While such rules are generally not neutral (as they depend explicitly on the underlying ordering of alternatives), some neutral rules can be characterized as augmented median voter rules. An example is dictatorship.

when viewed as examples of manipulable collective choice functions, a common logic explains them all.[13]

4.1 Amendment Agendas

In this section we briefly consider an institution that has received much attention in formal models of politics: the amendment agenda. Under an amendment agenda, alternatives are voted upon in an ordered sequence of pairwise votes. It is well-known that in the absence of a Condorcet winner, these agenda procedures are highly manipulable, with any alternative in the top cycle being attainable as a policy outcome depending on the sequence of votes taken. Thus, amendment agendas are, in general, not neutral, because they privilege alternatives appearing later on in the agenda.[14] It is also well-known that in the presence of a Condorcet winner, any sequence of voting will yield the Condorcet winner as an outcome, regardless of whether all individuals vote sincerely or all vote sophisticatedly. Thus, over the domain $\mathcal{S}^n$, amendment agendas *are* neutral: if the collection of ballots an amendment agenda is given is single-peaked, then so is any permutation of that collection, and the outcome of voting will be the Condorcet winner (and the permuted Condorcet winner) of each ballot profile.

What is perhaps less well-known is that amendment agendas are highly manipulable at sincere profiles of ballots, even in the presence of a Condorcet winner.[15] While, in the presence of a Condorcet winner, both fully sincere and fully strategic profiles of ballots yield the same outcome, these ballot profiles are not the same. Sincere profiles are manipulable; sophisticated profiles are not. To see this, consider the amendment agenda pictured in Figure 1, in which alternatives x and y are first put to a vote via majority rule, and

[13]The points made in this context are essentially extending arguments of Schofield (1995) and Austen-Smith and Banks (1998).

[14]The last alternative considered in a pairwise vote need only defeat the winning alternative that preceded it in order to become a policy outcome. However, an alternative considered first must defeat every other policy in order to be chosen as a policy outcome.

[15]Others have noted that when the behavioral assumption of sincerity or sophistication is not uniformly made across all voters, amendment agendas may no longer be Condorcet consistent. See Denzau et al. (1985) and Austen-Smith (1987), among others.

the winner is then pitted against z in order to determine the final outcome. Suppose that there are three individuals with the following preferences: xP_1yP_1z, yP_2zP_2x, and zP_3yP_3x. This preference profile is single peaked, and is pictured graphically in Figure 2; it yields y as a Condorcet winner.

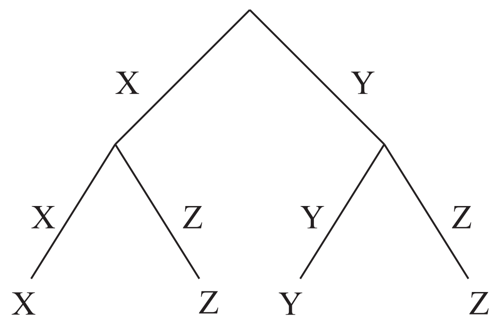

Figure 1: A two-stage amendment agenda

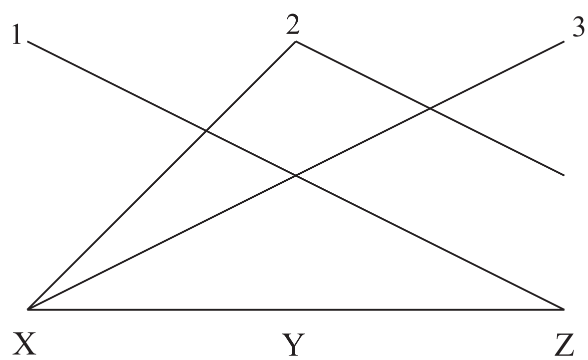

Figure 2: A single-peaked preference profile

Under a truthful collection of ballots, the amendment agenda pictured in Figure 1 yields y as an outcome: y defeats x at the first stage of voting by the votes of Players 2 and 3, and y defeats z at the second stage by Players 1 and 2. Now consider a collection of ballots in which Players 1 and 2 truthfully reveal their preferences over alternatives, but Player 3 claims to have the preference ordering $zP_3'xP_3'y$. Under this ballot profile our amendment agenda now yields z as the winner: x defeats y at Stage 1 by the (sincere) vote of Player 1 and the (insincere) vote of Player 2, and z defeats x at Stage 2 by the sincere votes of both Players 2 and 3. Furthermore, this is a beneficial manipulation by Player 3, as it enables him to attain his ideal point as the policy outcome.

Clearly the insincere ballot of Player 3 is not single-peaked with respect to the underlying ordering of alternatives. However, without *a priori* restricting how people can cast votes, manipulation is endemic to this form of agenda, even when the majority will is clearly well-defined. And we know of no real-world institution that restricts how pairwise votes may be cast. At the same time, when handed a truthful profile of ballots, the amendment agenda produces outcomes and sequences of votes that are consistent with pairwise majority voting. Thus, satisfying Arrow's conditions does not, "by easy implication" imply satisfaction of the conditions of Gibbard-Satterthwaite on single-peaked domains, as claimed by Dryzek and List (2002). Pairwise majority voting is transitive, weakly Paretian and IIA (and thus, neutral) when a collection of preferences is single-peaked, and produces outcomes consistent with those produced by an amendment agenda. Amendment agendas are not, however, strategy-proof.

4.2 The Romer-Rosenthal Setter Model

We have just shown that amendment agendas are manipulable even when preferences are single-peaked. However, the successful manipulation of an amendment agenda is only possible with a ballot that is not single peaked with respect to the true ordering over alternatives. In this section we provide an example of manipulation in which an insincere coalition is not required to break single-peakedness in its members' ballots in order to strictly benefit from insincerity.

Romer and Rosenthal's *setter model* considers a one-dimensional policy space in which an individual (the "setter") is chosen to propose a policy to the group, who then vote that policy up or down. If the policy is accepted, it is implemented. If not, the result is an exogenous status quo policy, x_Q. Romer and Rosenthal show that in this setting, the policy outcome is the setter's most-preferred policy that also defeats the status quo. Although a majority rule core will exist in this setting, the policy outcome need not be an element of the core.

For the purposes of this example, let $P(x) = \{y \in X : yP_ix, \forall i \in N\}$. Thus, $P(x)$ is the collection of policies that all players unanimously prefer to x. Let S represent the player who is the *setter*, and R represent the player who is the *receiver*, and let the ideal points of these players be denoted x_S and x_R. There are a number of ways that we could represent the outcomes generated by the two-player setter model as a collective choice function, and we choose the following: for any profile of ballots, $\rho = (P_S, P_R)$, $\phi(\rho) = x$ where $x \in P(x_Q)$ and xP_Sy for all $y \in P(x_Q) \setminus \{x\}$. Thus, outcome x is the setter's top-ranked policy that defeats the status quo.

Figure 3 depicts the spatial location of an exogenous status quo policy and the ideal points of both a setter and a receiver. The receiver's single-peaked preferences are also pictured, with the relevant information being that policy x is the setter's favorite policy that the receiver will agree to replace the status quo with. Thus, x is the outcome generated by the setter model. Clearly the receiver would be made strictly better off if the setter believed that he had preferences such as those pictured in Figure 4, which would yield a policy outcome of x' rather than x. In other words, the receiver *wishes* that the setter thought that he liked the status quo more than he actually does. If the receiver was able to credibly misrepresent his preferences to the setter, he would. Moreover, the receiver would not even have to alter his reported ideal point in order to strictly benefit from misrepresenting his preferences; he need only alter his relative preference for the status quo policy.

While this example is obviously quite simple, it demonstrates a fundamental susceptibility to manipula-

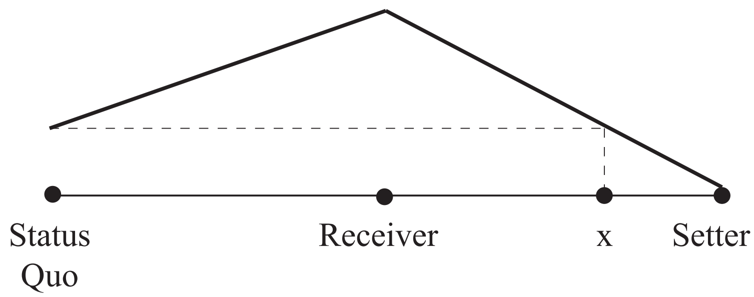

Figure 3: A two-player setter model

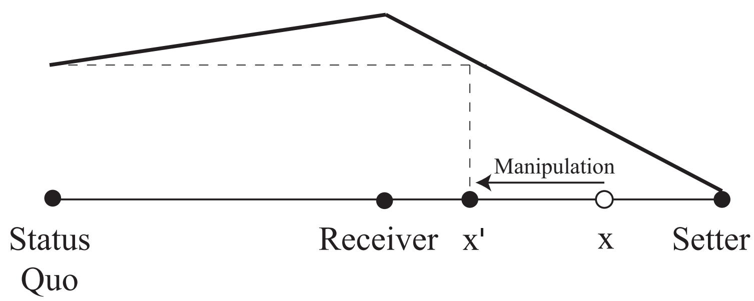

Figure 4: Successful manipulation by R with a single-peaked ballot

tion that many collective choice institutions with an agenda setter will have, even in settings where there is an unambiguous ordering to the alternatives under consideration that all ballots must adhere to (e.g. settings in which alternatives can be numerically ordered, such as tax rates). The problem is this: we know that in settings with an *a priori* ordering over alternatives, in which both preferences and ballots must adhere to single-peakedness with respect this ordering, every strategy-proof mechanism must be implementable as an augmented median voter rule. While a thorough description of these rules is beyond the scope of this paper, these rules necessarily take as an input only the top-ranked alternative of each individual's ballot. In other words, they must be invariant to how individuals rank all alternatives on their ballots, save for their topmost choices. This condition is known as *peaks only* and has been well-discussed in the literature on strategy-proof mechanisms on single-peaked domains with known orderings.[16]

The setter model, along with any mechanism in which a person or group targets a policy proposal for acceptance by a separate person or group, necessarily requires more information than peaks only. In particular, and as illustrated in the figures above, it requires information about how the status quo policy fares in relation to a new alternative under consideration, and under certain profiles this information cannot be obtained by simply knowing the ideal points of all players. In the formalization of this observation below, we assume that there at least two distinct coalitions that are blocking for (x, q) (i.e. can block alternative x with status quo q). One of these coalitions is a single individual, "setter" $i = L_S$ and the others are "receiving coalitions," with the collection of receiving coalitions denoted $\mathcal{L}_R$. Let $P(q;\rho) = \{x \in X$ such that $\nexists L \in \mathcal{L}_R$ with $qP_j x$ for all $j \in L\}$. In words, $P(q;\rho)$ is the collection of policies that cannot be blocked with status quo q by any receiving coalition submitting ballots as in ρ. The following collective choice function captures many institutional environments with an agenda setter.

Definition 13 (The setter collective choice function)**.** *For some* $i \in N$*, the* setter collective choice function *is* $\phi_S(\rho) = \text{Top}\ (P_i|_{P(q;\rho)})$.

Proposition 1. *The setter collective choice function is coalitionally manipulable at some profile* $\rho \in \mathcal{S}_Q^n$*. Furthermore, it is manipulable by an insincere ballot profile that is single-peaked with respect to* Q*.*

Proof: Let Q order the alternatives $x^1 < x^2 < ... < x^S$. Consider a profile $\rho \in \mathcal{S}_Q^n$ with $x_i^* = x^S$, and with $x_j^* = x^2$ for all $j \neq i$. Let $q = x^1$. If ρ is such that $x^S P_j q$ for all $j \neq i$, then $\phi_S(\rho) = x^S$. If ρ is such that $qP_j x^S$ for all $j \neq i$, then $\phi_S(\rho) \neq x^S$. Thus, there exist situations in which $\phi_S(\rho)$ cannot be determined without more information than the ideal points of all players; $\phi_S(\rho)$ is thus manipulable on $\mathcal{S}_Q^n$ with some

[16] In addition to Austen-Smith and Banks (2004, p. 42) who provide a proof of the necessity of the "peaks only" condition in these settings, see also Moulin (1980), among many others.

ballot $\rho' \in \mathcal{S}_Q^n$. (Austen-Smith and Banks, 2004, Lemma 2.5) $\square$

The above proof demonstrates that subgame perfect equilibrium outcomes cannot generally be represented as nonmanipulable "collective choice functions" of preferences, even in settings where preferences are single-peaked with respect to a known underlying ordering of alternatives. For example, when the ideal points of members of a receiving coalition lie strictly between the setter's ideal point and the status quo, it is impossible to deduce whether those individuals prefer the setter's ideal point to the status quo or vice versa without more information than ideal points. These institutional environments necessarily use more information than peaks only in order to produce collective outcomes, and thus, are manipulable.

4.3 The Ally Principle

Collective choice bodies frequently appoint delegates to make policy decisions on their behalf. This delegation underlies both representative democracy and administrative government, two rather notable innovations in government in the last few centuries. For example, legislatures routinely charge internal and external agents, such as committees, executive agencies, and courts, with carrying out various aspects of public policy. When policy-making power is delegated to an individual or group, a natural question is what kind of delegate should be selected. Most pointedly, is it the case that the legislature should select an agent or group of agents that will always pursue the policy goals that the legislature would itself have chosen? One answer to this question in a wide class of delegation problems falls under the label of the "ally principle," *e.g.*, "All else equal, a rational boss should choose her closest ally as an agent." (Bendor and Meirowitz, 2004, p. 300).

The arguments we presented in the previous section demonstrate a limitation of this reasoning: while it may not have been subgame perfect for the receiver in the setter model to misrepresent his preferences, what if the receiver, as a principal, was able to appoint a delegate to make decisions on his behalf? Unless the agent to whom power is delegated is a dictator, there is no reason to expect or recommend that a principal delegate authority to an agent whose preferences mirror his or her own.[17] For the agent to not be a dictator means, of course, that there are other political actors involved in the determination of the final policy choice. Our results imply that even when the principal's preferences and those of the other actors are single-peaked, this incentive to manipulate through the choice of an "insincere" or "unrepresentative" agent may still exist.

Considering the domain of single-peaked preferences implies that our results are directly applicable to the question of whether "the median voter within a legislature" might have an incentive to delegate gate-keeping power to an non-representative committee (*i.e.*, a committee of "preference outliers"). This point has been made recently within a canonical unidimensional spatial representation of bicameral bargaining in the U.S. Congress (Gailmard and Hammond, 2008). It has also recently reared its head within models of administrative policymaking (Boehmke et al., 2006) and civil service protections (Gailmard and Patty, 2007).

4.4 Institutional Choice: Is Pareto Efficiency Unambiguously Beneficial?

Maintaining the supposition of single-peaked preferences within a collective choice body is very convenient for the purpose of making the final point of this paper. First, our results indicate that the ally principle can fail even within the classical unidimensional spatial model. Secondly, Black's Median Voter Theorem tells us that majority preference is well-defined in this setting, implying that the failure of the ally principle (*i.e.*,

[17]Note that, in the baseline models examined by Bendor and Meirowitz (2004), the agent *is* a dictator. The principal exception to this is Extension 4b of their model. In this extension, the principal may revisit and revise the agent's decision *ex post*. Accordingly our results provide another demonstration of the logic behind why the ally principle is violated in this extension, as noted by Bendor and Meirowitz.

the incentive to manipulate the larger policymaking process) translates directly into a *collective* incentive to misrepresent "the legislature's preferences." Indeed, taking the argument a step farther, note that the collective incentives of such a legislative body can be, without loss of generality, presumed to be shared *unanimously* within the body in question. The following simple example illustrates this possibility in a precise fashion.

Example. Consider a situation in which three individuals, i, j and k, must collectively choose an individual to represent their interests in a bargain with two other individuals, a and b. The final choice will be determined through plurality rule, with each bargainer casting a vote for his or her favorite alternative. Any three-way tie is broken in favor of a. By Theorem 6, this collective choice function is coalitionally manipulable on $\mathcal{S}^3$, as it takes on full range within a set of three alternatives and is nondictatorial. Furthermore, it is individually manipulable on $\mathcal{S}^3$. The (single-peaked) preferences of the individuals are as follows:

	ρ
a	$x \succ y \succ z$
b	$y \succ x \succ z$
i, j, and k	$z \succ y \succ x$

It is clear that i, j, and k have an incentive to choose a representative with preferences $y \succ z \succ x$ or $y \succ x \succ z$. However, any weakly Paretian preference rule used to aggregate the preferences of i, j and k must return the ordering $z \succ y \succ x$. It follows that if the representative of i, j and k must interact with other players in a larger bargaining game, i, j, and k may actually prefer their representative to bargain for a Pareto-inferior alternative. $\square$

Notice that, in the example, every member of the constituency $\{i, j, k\}$ is made unambiguously better off by the selection of an apparently Pareto-inferior delegate or "intra-group" institution.[18] Accordingly, the logic behind the example has a direct implication for one of Arrow's axioms. In particular, the example demonstrates that when a collective choice rule is manipulable by a player $i \in N$ at a preference profile ρ, and when that "player's" preferences are the result of the aggregation of the preferences of a set of individuals, there will be situations in which those individuals will have a strategic incentive to aggregate their own preferences through an institution that is not weakly Paretian. The implications of this example can be taken to the applied analysis of institutions, as in Crombez et al. (2006), who argue that "gatekeeping" institutions are Pareto-dominated by "veto institutions" in terms of the intra-chamber equilibrium outcomes produced. The example suggests that, insofar as a collective choice body is part of a larger policymaking process, there will exist situations in which the members of the chamber in question can make themselves better off within that larger policymaking process by delegating gatekeeping authority to a (non-representative) individual or group that will effectively withhold or block one or more alternatives from the body – even alternatives that are *unanimously* preferred *by* the body to the status quo policy. In other words, gatekeeping institutions may not be so "deeply flawed from the perspective of social efficiency" when the body in question is interacting with other agents within a non-dictatorial collective choice institution (Crombez et al., 2006, p. 326).

5 Conclusions

Theorems 4, 5, and 6 imply that one must be careful in interpreting collective will in any real-world policymaking institution *even when preferences are presumed to be single-peaked*. This point is highly relevant

[18]In particular, we conceive of an "intra-group" institution as a mechanism by which individuals i, j and k choose the collective preference they wish to reveal to the other players in a larger policymaking process.

for those scholars who insist that majority rule cycles are infrequent or untroubling (*e.g.*, Mackie (2003)). Specifically, appeals to aggregate outcomes as indicators of collective will are not necessarily well-founded even when the majority will is assumed to exist. “Faithful representation” of the majority will within non-dictatorial institutions will occasionally take an insincere form. Accordingly, the normative, prescriptive, descriptive, and inferential issues raised by Arrow’s theorem and the Gibbard-Satterthwaite theorem are more than simple mathematical curiosities dreamed up for the purpose of scholarly debate.

Single-peakedness does not solve problems of cycling in the real-world because policymaking institutions are generally not neutral. Specifically, the presumption that individuals have single-peaked preferences is not sufficient to assume that the result of their aggregation is a well-defined collective will. Single-peakedness does not eliminate the possibility of gains through strategic manipulation within real-world institutions, because few (if any) policymaking institutions are dictatorial. Indeed, in the real world, the collective choices of any collective choice body on policy outcomes is necessarily conditioned on the actions of other actors who also participate in the policy process. An institution’s sovereignty over its *actions* must not be confused with sovereignty over *policy outcomes*. To the degree that political actors within the institution care about outcomes, the presumption that their individual and collective choices should be sincere in order to be considered faithfully representative is unfounded.

References

Arrow, K. J. (1951). *Social Choice and Individual Values*. John Wiley and Sons, New York, NY.

Austen-Smith, D. (1987). Sophisticated sincerity: Voting over endogenous agendas. *American Political Science Review*, 81(4):1323–1330.

Austen-Smith, D. and Banks, J. S. (1998). Social choice theory, game theory, and positive political theory. *Annual Review of Political Science*, 1(1):259–287.

Austen-Smith, D. and Banks, J. S. (2004). *Positive Political Theory II: Strategy & Structure*. University of Michigan Press, Ann Arbor, MI.

Ballester, M. A. and Haeringer, G. (2007). A Characterization of the Single-Peaked Domain. Working paper, Universitat Autònoma de Barcelona.

Bendor, J. and Meirowitz, A. (2004). Spatial Models of Delegation. *American Political Science Review*, 98(2):293–310.

Black, D. (1948). On the Rationale of Group Decision-making. *Journal of Political Economy*, 56:23–34.

Blin, J. M. and Satterthwaite, M. A. (1976). Strategy-proofness and single-peakedness. *Public Choice*, 26(1):51–58.

Boehmke, F., Gailmard, S., and Patty, J. (2006). Whose Ear to Bend? Information Sources and Venue Choice in Policy Making. *Quarterly Journal of Political Science*, 1(2):139–169.

Crombez, C., Groseclose, T., and Krehbiel, K. (2006). Gatekeeping. *Journal of Politics*, 68(2).

Davis, O., Hinich, M., and Ordeshook, P. (1970). An Expository Development of a Mathematical Model of the Electoral Process. *American Political Science Review*, 54:426–448.

Denzau, A., Riker, W., and Shepsle, K. (1985). Farquharson and Fenno: Sophisticated Voting and Home Style. *American Political Science Review*, 79(4):1117–1134.

Dowding, K. and Van Hees, M. (2007). In Praise of Manipulation. *British Journal of Political Science*, 38(1):1–15.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper and Row, New York.

Dryzek, J. and List, C. (2002). Social Choice Theory and Deliberative Democracy: A Reconciliation. *British Journal of Political Science*, 33(1):1–28.

Dummett, M. and Farquharson, R. (1961). Stability in Voting. *Econometrica*, 29(1):33–43.

Feld, S. and Grofman, B. (1988). Ideological Consistency as a Collective Phenomenon. *American Political Science Review*, 82(3):773–788.

Gailmard, S. and Hammond, T. (2008). Intercameral Bargaining and Intracameral Organization in Legislatures. Mimeo, Northwestern University.

Gailmard, S., Patty, J., and Penn, E. M. (2008). Arrow's theorem on single-peaked domains. *Mimeo*.

Gailmard, S. and Patty, J. W. (2007). Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise. *American Journal of Political Science*, 51(4).

Gibbard, A. (1973). Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601.

Krehbiel, K. (1998). *Pivotal Politics: A Theory of U.S. Lawmaking*. University of Chicago Press, Chicago, IL.

Mackie, G. (2003). *Democracy Defended*. Cambridge University Press, New York, NY.

McCubbins, M., Noll, R., and Weingast, B. (1994). Legislative Intent: The Use of Positive Political Theory in Statutory Interpretation. *Journal of Law and Contemporary Problems*, 57(1):3–37.

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

Moulin, H. (1980). On Strategy-Proofness and Single Peakedness. *Public Choice*, 35(4):437–455.

Pattanaik, P. (1976). Collective rationality and strategy-proofness of group decision rules. *Theory and Decision*, 7(3):191–203.

Poole, K. and Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll-Call Voting*. Oxford University Press, New York, NY.

Reny, P. J. (2000). Arrow's theorem and the gibbard-satterthwaite theorem: A unified approach. *Mimeo*, Department of Economics, University of Chicago.

Riker, W. (1992). The Justification of Bicameralism. *International Political Science Review/ Revue internationale de science politique*, 13(1):101–116.

Romer, T. and Rosenthal, H. (1978). Political Resource Allocation, Controlled Agendas, and the Status Quo. *Public Choice*, 33:27–43.

Satterthwaite, M. A. (1975). Strategy-Proofness and Arrows Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory*, 10:187–217.

Schofield, N. (1995). Coalition Politics: A Formal Model and Empirical Analysis. *Journal of Theoretical Politics*, 7:245–281.

Sen, A. K. (1966). A Possibility Theorem on Majority Decisions. *Econometrica*, 34(2):491–499.

Sen, A. K. and Pattanaik, P. K. (1969). Necessary and Sufficient Conditions for Rational Choice Under Majority Decision. *Journal of Economic Theory*, 1(2):178–202.

Ubeda, L. (2003). Neutrality in arrow and other impossibility theorems. *Economic Theory*, 23(1):195.

Young, H. P. (1995). Optimal Voting Rules. *Journal of Economic Perspectives*, 9(1):51–64.