

# Choosing How to Choose: Self-Stable Majority Rules and Constitutions

Salvador Barbera and Matthew O. Jackson \*

August 2000  
Revised: August 26, 2003

## Abstract

Constitutional arrangements affect the decisions made by a society. We study how this effect leads to preferences of citizens over constitutions; and ultimately how this has a feedback that determines which constitutions can survive in a given society. Constitutions are stylized here, to consist of a voting rule for ordinary business and possibly different voting rule for making changes to the constitution. We define an equilibrium notion for constitutions, called self-stability, whereby under the rules of a self-stable constitution, the society would not vote to change the constitution. We argue that only self-stable constitutions will endure. We prove that self-stable constitutions always exist, but that most constitutions (even very prominent ones) may not be self-stable for some societies. We show that constitutions where the voting rule used to amend the constitution is the same as the voting rule used for ordinary business are dangerously simplistic, and there are (many) societies for which no such constitution is self-stable rule. We conclude with a characterization of the set of self-stable constitutions that use majority rule for ordinary business.

---

\*Barbera is at CODE, Departament d'Economia i d'Historia Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain, (Salvador.Barbera@uab.es) and Jackson is at the division of Humanities and Social Sciences, 228-77, California Institute of Technology, Pasadena, California 91125, USA, (jacksonm@hss.caltech.edu). We thank Randy Calvert, Dan Levin, and Jean Mercier Ythier for very helpful discussions of the paper. We also thank Danilo Coelho, Gabrielle Demange, Anke Gerber, Annik Laruelle, Eric Maskin, Charlie Plott, Cheng-Zhong Qin, Jim Snyder, Guofu Tan, Gordon Tullock, Federico Valenciano, and Peyton Young for helpful conversations and comments, and Takehiko Yamato for detailed comments and suggestions on an earlier draft. We are also grateful for suggestions by Alberto Alesina and two anonymous referees that have improved the paper. Financial support under NSF grant SES-9986190 and under DCGYT Direcció General de Recerca projects PD-98-0870 and SGR-980062 is gratefully acknowledged. This project was initiated during a visit of Barbera to Caltech.

# 1 Introduction

Different societies use different rules to make collective choices. Moreover, societies often use different rules to decide on different types of issues. We offer a theory that accounts for this diversity. Specifically, we show why some constitutions will be stable over time, while others will not, and we also show how this relates to the structure of the society.

Societies realize that change is inevitable, and they usually (but not always) build rules for amendment into their constitutions. From this point of view, it is clear that a constitution is a choice of a society, and a choice that may be revisited over time. As such, in order to be able to make any meaningful predictions about the types of constitutions that we should expect to see in the world, we need a theory about which constitutions will survive over time, given that they are subject under their own rules to amendments.

Thus, simply put, in this paper we study the constitutions that can be considered as equilibrium constitutions. More explicitly, we name these “self-stable” constitutions, to emphasize the idea that change is governed by the constitution itself.

It is important to emphasize that our goal is not to say which constitutions are better than others, nor to explain why a society allows for amendments or flexibility, nor why a society picks one constitution over another. Our goal is simply to provide a theory of which constitutions can survive over time, and which ones cannot. With this said, there are types of constitutions that are particularly prominent in the world (for instance, involving majority rule) that are worthy of special attention in understanding when they are self-stable, and when they are not.

Our study begins with the most basic form of constitution that one could imagine. It is simply a specification of a voting rule. The idea is that under such a constitution, any decisions that the society will make, including voting over changes to the constitution, will be governed by that voting rule. There we identify sufficient conditions for such simple constitutions to be self-stable, but also find that there are conditions under which no such constitution is self-stable.

This then turns our attention to a slightly more complicated form constitution, that might be thought of as a stylized version of what one sees in the world. The idea is that the constitution specifies a voting rule for passing new legislation, except when it comes to “special” types of proposals. For such special proposals - in particular amendments to the constitution - a different voting rule is used. For example, many societies take decisions by majority rule, but require a two-thirds majority to amend the constitution. For instance, the U.S. Senate uses majority rule, and a 67/100 rule to change the senate rules. In fact, under the filibusters that are possible in the senate, one needs 60/100 votes to call a vote and so the effective voting rule might be thought of as a (60%,67%) constitution rather than a (50%,67%) constitution. An interesting (unstable) example arose recently in California. Under the law until 2000, school bond and tax issues required a 2/3 majority of the participating voters to pass. So we might think of these votes as having the basic voting rule be 2/3 of the voters. However, propositions (initiatives that may be placed on the ballot through a variety of means) in California may be passed with a 1/2 majority. In particular,

one can place a proposition on the ballot which changes the vote required on such issues. Thus, one can amend the state's voting rule by a 1/2 vote. In fact in the 2000 election, Proposition 39 suggested changing the voting rule on school bond and tax issues from 2/3 to 55%. Interestingly, Proposition 39 passed with 53.4% (as reported by the Secretary of State of California) of the vote. Having a 2/3 majority voting rule that can be amended by a 1/2 vote is inherently unstable.

Our analysis shows that some of these more complicated constitutions that involve separate voting rules for standard decisions and for amendments, are always self-stable. But we also find that even very natural constitutions, like the one-half, two-thirds described above, may fail to be self-stable under some circumstances. More importantly, we determine exactly how self-stability hinges on the preferences over rules that are held by the citizens, and we show that these preferences are endogenously determined by each citizen's assessment of his or her relative position in the political spectrum.

To insist on the relevance of our analysis, we emphasize that special majorities are often required by constitutions on many different issues, and not only in order to make amendments to the constitution. For example, the Council of the European Union makes decisions by qualified majority on an increasing list of subjects. Our theory also explains why, for any given family of issues, qualified majorities may be desired by voters, whether or not sustained by further constitutional arrangements. Because of that, it is also interesting to study the stability of qualified majority rules, for the case where no special provisions are made to change them. A striking example of a situation when the standard rule is explicitly stated to be the one required for rule change is provided by the draft Constitution proposed by the Convention of the EU. The second paragraph of article 24.4 reads as follows: "Where the Constitution provides in Part III for the Council of Ministers to act unanimously in a given area, the European Council can adopt, on its own initiative and by unanimity, a European decision allowing the Council of Ministers to act by qualified majority in that area".

We are aware that any simple explanation of a phenomenon as complex as the choice of a constitution can only be partial. Many factors other than self-stability influence the choice and the persistence of constitutional arrangements. Nevertheless, self-stability is a central property that one needs to understand in order to develop a robust theory of constitutions. It is important to reflect on this property, which seems to have passed unnoticed in the literature, and yet formalizes an equilibrium requirement that one should expect to be satisfied by any persistent set of rules.

## The Related Literature

Constitutional design and properties of voting rules are topics that have been extensively studied in political science and social choice theory, dating to the classics, such as Rousseau (1762) who explicitly discussed how the size of a majority required in a voting rule should be related to the importance of the question at hand.<sup>1</sup> Buchanan and Tullock (1962) were the

---

<sup>1</sup>Some recent references from the large literature that relates to issues regarding majority size includes Caplin and Nalebuff (1988), Austen-Smith and Banks (1997), Feddersen and Pesendorfer (1998), and Dasgupta and Maskin (1998).

first to raise the issue of choosing how to choose, and they raise it explicitly in the context of constitutional design. To quote from their work (page 6):

“When we recognize that “constitutional” decisions themselves, which are necessarily collective, may also be reached under any of several decision-making rules, the same issue is confronted all over again. Moreover, in postulating a decision making rule for constitutional choices, we face the same problem when we ask: How is the rule itself chosen?”

While Buchanan and Tullock raise the issue of choosing how to choose, they end up stepping around it and instead focussing on the role of full consent in decision making, including decisions regarding constitutional choice. In contrast, the approach that we take here addresses this problem of choosing constitutions head-on.

Recent research on constitutional structure (for instance, Persson and Tabellini (2000), Persson (2002), and Aghion, Alesina and Trebbi (2002)) advances both theoretical arguments and empirical tests stressing the importance of constitutional arrangements for economic performance, as well as feedback from the economy on political institutions. Our work is wholly theoretical analysis and one with a complementary focus, namely on understanding which political institutions will be stable. As emphasized by Persson in his EEA presidential address (Stockholm, August 2003), there is room and need for both theoretical and empirical approaches in the long road towards a full grasp of these interactions.

The model of voter uncertainty that we work with was first proposed in the early nineteen seventies in a series of brilliant papers (most of which are collected in a volume edited by Niemi and Weisberg (1972), and discussed in what follows), inspired by a seminal work of Rae (1969), whose purpose was to justify the use of majority rule. Some of our results reinforce the idea that majority rule is special.<sup>2</sup> In particular, it emerges as the natural rule for day-to-day decisions in self-stable constitutions, coupled with an adequate supermajority for the change of rules (see Theorems 4 and 5). Yet, our analysis also clarifies that other majority sizes may be self-stable in cases where simple majority would not be.

Koray (2000) is an important predecessor to ours on the subject of a choice of voting rules. Koray outlines a method for viewing social choice functions themselves as alternatives, so that one can ask whether a social choice function always selects itself. He shows that given enough richness of preferences the only self-selective social choice functions are dictatorial. This differs in key ways from our self-stability and impossibility results are not an issue in our analysis. We end up with dramatic differences in the model and the results, so that the only real tie between our study and Koray’s is in the common interest of endogenizing the way in which societies make choices.<sup>3</sup>

---

<sup>2</sup>Due to its salience, majority rule has been analyzed and justified from very different angles. Condorcet (1785) provided a classical justification for its use through what is now called the Condorcet Jury Theorem (see also Young and Levenglick (1978)). Another type of justification comes from axiomatic analysis (May (1952)). Our approach differs from the axiomatic, because we treat the decision rules as choice variables, and from the Jury Theorem approach because our voters may have conflicting objectives.

<sup>3</sup>Here are some of the main differences. First, our concept of self-stability only requires that a voting rule should not be beaten by another rule when the given rule is used, which is different from saying that a rule must select itself. Another way to say this is that in our setting there is a special standing to the status-quo

We finish by mentioning some recent work directly related to voting over voting rules and to our approach. A paper by Sosnowska (2002) extends the model that we develop here to consider weighted voting rules, a paper by Wakayama (2002) extends the model to allow for abstention, and Coelho (2002) considers notions of maximin in place of self-stability. Maggi and Morelli (2003) study such voting in international organizations. Polborn and Messner (2002) consider the choice of voting rules and self-stability in the context of an overlapping generations models where differences in cohort size and voting rules affect the passage of reforms that involve costs and delayed benefits.

## 2 Definitions

### Voters and Alternatives

$N = \{1, \dots, n\}$  is a set of *voters*.

The voters will face votes over pairs *alternatives*. We denote the terms of these pairwise choices as  $a$  and  $b$ . Alternative  $a$  is interpreted as the *status-quo*. Alternative  $b$  is interpreted as a *change*.

### Voting Rules

Each voter casts a vote in  $\{a, b\}$ .

A *voting rule* is characterized by a number  $s \in \{1, \dots, n\}$ .<sup>4</sup> If at least  $s$  voters say “ $b$ ” then  $b$  is elected, and  $a$  is elected otherwise.

Some examples of voting rules are as follows.

If  $s = 1$ , then  $b$  is elected whenever there is at least one voter for change, and so  $a$  is elected only when it is unanimously supported.

If  $s = n$ , then  $b$  is elected if there is unanimous support for change, and  $a$  is elected as soon as at least one a voter supports it.

If  $n$  is odd and  $s = \frac{n+1}{2}$  or  $n$  is even and  $s = \frac{n}{2} + 1$ , then the voting rule is the standard *majority rule*.<sup>5</sup>

As majority rule is referred to at several points in what follows we denote it by  $s^{\text{maj}}$ . Thus,  $s^{\text{maj}} = \frac{n+1}{2}$  if  $n$  is odd and  $s^{\text{maj}} = \frac{n}{2} + 1$  if  $n$  is even.

---

alternative, which can provide an asymmetry not present in the more abstract social choice setting analyzed by Koray. Second, the underlying setting here considers votes over two (possibly uncertain) alternatives at a time, rather than making selections from three or more (known) alternatives.

<sup>4</sup>Allowing for  $s = 0$  or  $s = n + 1$  results in degenerate voting rules that always choose  $b$  or always choose  $a$ , respectively. We focus on rules where there is a real choice to be made.

<sup>5</sup>When  $n$  is even, there are two possible choices:  $\frac{n}{2}$  and  $\frac{n}{2} + 1$  depending on which alternative wins in the case of a tie. For simplicity, we break ties in favor of the status quo in this case. None of the analysis that follows is dependent on tie-breaking conventions.

Note that our definition of a voting rule presumes anonymity. We discuss this property in the concluding remarks.

## Voter Preferences

Voters have preferences over voting rules, as the voting rule will affect the future of the society. Let voter  $i$ 's preferences over voting rules be represented by the utility function  $U_i : \{1, \dots, n\} \rightarrow \mathbb{R}$  where  $U_i(s)$  represents voter  $i$ 's utility for voting rule  $s$ .

In the next section, we analyze voters' preferences in detail. For the purposes of introducing our definitions of self-stability, it is sufficient simply to know that voters have preferences over voting rules.

## Self-Stable Voting Rules

A voting rule  $s$  is *self-stable* (for society  $p$ ) if  $\#\{i \mid U_i(s') > U_i(s)\} < s$  for every  $s' \neq s$ .

The property of self-stability ensures that a given voting rule would be robust to change if used for making decisions.

We should emphasize that self-stability may be thought of as an equilibrium concept. As with many equilibrium concepts, we do not model how one reaches equilibrium, nor do we model how the world might select among equilibria if there are several. What we can say is that a self-stable rule would stay in place if reached, while other rules would tend not to.

## Self-Stable Constitutions

We also explore the consequences of admitting constitutions that allow for different voting rules to be used for making different types of decisions. A constitution can specify one voting rule to be used on all issues except for the change of this voting rule, where a different rule may be used.

A *constitution* is a pair of voting rules  $(s, S)$ , where  $s$  is to be used in votes over the issues  $a, b$  and  $S$  is to be used in any votes regarding changes from  $s$  to any other rule  $s'$ .<sup>6</sup>

A constitution  $(s, S)$  is *self-stable* if  $\#\{i \mid U_i(s') > U_i(s)\} < S$  for any  $s'$ .

Self-stability of a constitution requires that the preferences of voters be such that there does not exist a voting rule  $s'$  that would defeat the constitution's prescribed voting rule  $s$  to be used for choices over issues, when these two voting rules are compared under the constitution's voting rule  $S$ , to be used for choices over rules. So, a self-stable constitution is one that would not be changed once in place.

The main focus of this paper is to say something about which voting rules and constitutions are self-stable. The idea is that these are the only rules that will survive in the long run in a society, and so it makes sense to understand what they look like.

---

<sup>6</sup>As pointed out to us by Randy Calvert, one could also think of a more general nesting of rules, where one thinks of a voting rule  $S'$  to amend  $(s, S)$ , and so on; and it might be interesting to consider when these may be truncated (as effectively the case of a pair means that the same  $S$  is used for all higher orders).

### 3 Induced Preferences over Voting Rules

In order to say something about self-stability, it is important to understand the structure of voters' preferences over voting rules.

#### Timing and Uncertainty

Let us consider a two period world. As we argue shortly, this easily extends to an infinite horizon model.

In period 2, a vote will be taken over two decisions  $a$  and  $b$ . At this time, each voter knows his or her preferences over  $a$  and  $b$ .

In period 1, voters do not yet know their preferences over  $a$  and  $b$ . A voter can be characterized by a probability  $p^i \in (0, 1)$ , that he or she will prefer  $b$  to  $a$  at the time of the vote.<sup>7</sup>

The realizations of voters' support for the alternatives are independent. For instance, the probability that voters 1 and 2 support  $b$  while voter 3 supports  $a$  is  $p_1 p_2 (1 - p_3)$ .

A voter gets utility 1 if his preferred alternative is chosen in the vote, and utility 0 otherwise.<sup>8</sup>

This sort of uncertainty was first considered in Badger (1972) and Curtis (1972), and we will make use of some of their results about voter preferences in what follows.

The *society* of voters is represented by a set of voters  $N$  and a vector  $p = (p_1, \dots, p_n)$ .

In what follows, we treat the society  $(N, p)$  as given and so will often suppress the fact that preferences will depend on these parameters, except where we want to specifically point out this dependence.

#### The Timing of Voting

In this world, a vote over the alternatives  $a$  and  $b$  will take place in period 2. There are two different times at which a vote over voting rules could be taken and conceivably be relevant. The first is in period 1 where voters do not yet know their preferences over the alternatives (but know the  $p_i$ 's). The second is in period 2, just before the vote over the alternatives, at a time where voters know which alternatives they support.<sup>9</sup>

The only votes over voting rules that are of any interest turn out to be in period 1, as votes over voting rules in period 2 are of no consequence. This is easily seen as follows. Suppose that the voting rule is  $s$  at the beginning of period 2. Let  $x$  be the number of voters

---

<sup>7</sup>Extensions to the case where  $p_i$  can be 0 or 1 are straightforward. These cases complicate some of the calculations and proofs when we divide by  $p_i$  or  $1 - p_i$ , but are still easily directly handled as special cases. To keep an uncluttered exposition, we leave the cases where  $p_i = 0$  or 1 to the interested reader.

<sup>8</sup>This presumes that a voter cares as much for getting change when preferring change over the status quo, as the voter cares for preserving the status-quo when preferring the status quo over change. We discuss the role of this assumption in detail in the concluding remarks.

<sup>9</sup>One could also conceive of voting over voting rules at some time 0, say "behind the veil of ignorance" and before the  $p_i$ 's are known. This might set a starting point for the evolution of the voting rule, but the only rules that would survive past period 1 would be self-stable ones. And so, that is our focus.

who support  $a$ , and  $n - x$  be the number of voters who support  $b$ . If  $n - x \geq s$ , then  $b$  will pass under voting rule  $s$ . In this case, these  $n - x$  voters will be happy with the voting rule  $s$  and would want to change it to any voting rule that would lead  $b$  not to pass. Since there are  $n - x \geq s$  such voters, no voting rule that could make a difference could defeat  $s$ . Next, consider the case where  $n - x < s$ . In this case, the  $n - x$  voters who prefer  $b$  would like to lower the voting quota to some  $s' < s$ , so they could get  $b$  to pass. However, the remaining  $x$  voters would prefer to keep  $s$  as it is, because they prefer  $a$  to  $b$ . Thus, these voters would vote against any such change, and again the voting rule would not be changed in any way that could make a difference.

Thus, we have argued that *the only interesting votes over voting rules have to come at a time where voters are still uncertain about their preferences over alternatives*. Therefore, in what follows, we analyze the preferences and votes over voting rules at period 1, when voters know their  $p_i$ 's but do not yet know their realized preferences over the alternatives.

### Induced Preferences over Voting Rules

Given the likelihood of different patterns of support for  $a$  and  $b$ , a voter can calculate his or her expected utility (at time period 1) under each voting rule  $s$ . Let  $U_i(s)$  be the expected utility of voter  $i$  if voting rule  $s$  is used. This is expressed as follows. For any  $k \in \{0, \dots, n - 1\}$ , let  $P_i(k)$  denote the probability that exactly  $k$  of the individuals in  $N \setminus \{i\}$  support the change. We can write

$$P_i(k) = \sum_{C \subset N \setminus \{i\} : |C|=k} \times_{j \in C} p_j \times_{\ell \notin C} (1 - p_\ell). \quad (1)$$

and

$$U_i(s) = p_i \sum_{k=s-1}^{n-1} P_i(k) + (1 - p_i) \sum_{k=0}^{s-1} P_i(k). \quad (2)$$

### Single Peaked Preferences

The usual definition of single-peaked preferences requires that all alternatives can be ranked from left to right, that one alternative  $\hat{s}$  is best, and that the alternatives that one encounters by moving leftward (or rightward) away from  $\hat{s}$  are considered worse and worse. Our definition here will be slightly weaker, as it allows a voter to have two peaks.<sup>10</sup> In particular, it is possible that  $U_i(\hat{s}) = U_i(\hat{s} - 1)$ . For instance, in a society where  $n$  is even and each  $p_i = p$  for all  $i$ , all individuals will be indifferent between  $n/2$  and  $n/2 + 1$ .

$U_i$  is *single-peaked* if there exists  $\hat{s} \in \{1, \dots, n\}$  with  $U_i(\hat{s}) \geq U_i(s)$  for all  $s \in \{1, \dots, n\}$  such that  $U_i(s) > U_i(s - 1)$  for any  $\hat{s} > s > 1$  and  $U_i(s - 1) > U_i(s)$  for any  $n \geq s > \hat{s}$ .

Let  $\hat{s}_i$  denote the *peak* of voter  $i$ .

---

<sup>10</sup>These could be referred to as single-plateaued preferences following the literature. However, given that such indifference can only occur between two points and happens non-generically (in  $p$ ) we stick with the term single-peaked.



In the case where a voter has twin-peaks, the definition above selects the higher of the two peaks as  $\hat{s}_i$ . This is simply a convention and does not matter in any of the results that follow.

The following result is due to Badger (1972). We include a proof in the appendix, for completeness.

**LEMMA 1** [Badger (1972)] *For any society (profile of  $p_i$ 's), every voter's preferences over voting rules are single-peaked.*

The following example gives some insight into voters' preferences over voting rules.

**EXAMPLE 1** *Single-Peaked Preferences*

Let us consider a simple society where agents can be divided into two different groups,  $N^1 = \{1, \dots, 4\}$  and  $N^2 = \{5, \dots, 10\}$ , where the  $p_i$  of each voter  $i$  in group 1 is  $p^1 = .01$  and in group 2 is  $p^2 = .99$ .

In this society, the corresponding peaks of preferences over voting rules are  $\hat{s}^1 = 8$  and  $\hat{s}^2 = 4$ .

Let us examine why  $\hat{s}^2 = 4$ , as this will help us to understand preferences more generally. This can be verified by direct calculations, but also can be seen in an intuitive manner. Let us consider a voter in  $N^2$ . Consider a scenario where exactly three voters end up supporting change. Given the extreme values of  $p^1 = .01$  and  $p^2 = .99$ , if there are three voters who end up supporting change, it is very likely that all of those voters are from  $N^2$ . Given that there are six voters in  $N^2$  this leads to a probability of nearly  $1/2$  that a voter in  $N^2$  would assign to supporting change conditional on three voters supporting change. Although this probability is nearly  $1/2$ , it is still less than  $1/2$  due to the small probability that some of the voters in  $N^1$  will be among those supporting change. So, a voter in  $N^2$  will prefer that society choose the status quo conditional on three voters supporting change. If we consider a scenario where exactly four voters end up supporting change, then the conditional probability that a voter in  $N^2$  would assign to being one of the supporters of change is nearly  $2/3$ . Since it is above  $1/2$ , a voter in  $N^2$  will prefer that society choose change conditional on four voters supporting change. Given these two observations it follows that  $\hat{s}_2 = 4$ . Similar reasoning leads to  $\hat{s}_1 = 8$ .

Generally we can think of a voter considering each possible scenario of numbers of supporters for each of the alternatives. For each scenario the voter determines which group they are more likely to fall in. The voter's most preferred voting rule ( $\hat{s}_i$ ) corresponds to the scenario with the smallest sized group supporting change for which the voter finds it more likely that he or she will support change. We can see that if the voting rule is raised or lowered from 4, then there will be some scenarios where the choice will be made in favor of the group that the voter finds it less likely that he or she will fall in. This is the explanation for why we see single-peaked preferences. We can also see why it is rare for a voter to have

twin peaks - as that can only happen in a case where the voter assigns probability of exactly 1/2 to each of the two groups in some scenario.

### Intermediate Preferences and Single Crossing

While Lemma 1 tells us that each voter's preferences over voting rules have the nice property of single-peakedness, the following lemma tells us about how different voters' preferences are related to each other. There are two properties that are useful in noting.

A society of voters has preferences satisfying the *single crossing property* if for any  $i$  and  $j$  with  $p_j \geq p_i$ ,

$$U_i(s) - U_i(s') \geq U_j(s) - U_j(s')$$

for all  $s \geq s'$ .

As we shall see, the single crossing property is satisfied in this model. The single crossing property allows us to order preferences over voting rules in terms of the  $p_i$ 's; but more importantly also implies that the preferences are intermediate.

A society of voters has *intermediate preferences* if for any  $i, j, k$  with  $p_j \geq p_k \geq p_i$ :

- $U_i(s) \geq U_i(s')$  and  $U_j(s) \geq U_j(s')$  imply that  $U_k(s) \geq U_k(s')$ , and
- $U_i(s) > U_i(s')$  and  $U_j(s) > U_j(s')$  imply that  $U_k(s) > U_k(s')$ .

Intermediate preferences are usually defined by requiring that there exists some ordering over individuals so that when two individuals have the same ranking over two alternatives, then individuals between them in the ordering have that same ranking (e.g., see Grandmont (1978)). Here the natural ordering over individuals is in terms of their  $p_i$ 's, the distinguishing characteristic of voters, and so we take the shortcut of defining intermediate preferences directly in terms of that ordering. Hence, a society will have intermediate preferences over voting rules if whenever two voters with  $p_i$  and  $p_j$  agree on how to rank two rules  $s$  and  $s'$ , then all voters with probabilities  $p_k$  between  $p_i$  and  $p_j$  will also agree on the way to rank these two rules. The simple model we are considering has the following strong feature.

**LEMMA 2** *Every society has preferences over voting rules that satisfy the single crossing property and are intermediate.*

The proof of Lemma 2 appears in the appendix. The intuition for why the voters' peaks over voting rules follow an inverse order to the voters  $p_i$ 's (Corollary 1 below), is fairly straightforward, as voters with higher  $p_i$ 's are more likely to favor change and thus will be in favor of a lower quota than voters who are less likely to favor change. While ordering the peaks is intuitive and useful, we emphasize that Lemma 2 has much stronger implications, as it relates preferences over arbitrary values of  $s$  and  $s'$ , including those falling on opposite sides of a set of voters' peaks. This additional structure will also be useful in what follows.

The proof of these aspects of preferences builds inductively from preferences over adjacent voting rules, and involves direct comparison of the expressions of differences in expected utilities for different voters. Details are in the appendix.

As just mentioned above, Lemma 2 has the following useful corollary (see the proof of Lemma 2).

**COROLLARY 1** *For any society,  $\hat{s}_i \geq \hat{s}_j$  whenever  $p_j \geq p_i$ .*

There are some other facts about the location of the voters' peaks that are worth emphasizing. The *relative* ordering of  $p_i$ 's is not only important in determining the relative ordering over the  $\hat{s}_i$ 's, but it is also critical in determining the actual values of the  $\hat{s}_i$ 's. This is seen in the following lemma, which states that *regardless* of  $p$ , there is always some voter who has a peak at least as high as  $s^{\text{maj}}$  and some other voter who has a peak no higher than  $s^{\text{maj}}$ .

**LEMMA 3** *For any society there exist  $i$  and  $j$  such that  $\hat{s}_i \geq s^{\text{maj}} \geq \hat{s}_j$ .*

The proof of Lemma 3 is based on the following reasoning. The unique maximizer of  $\sum_i U_i(s)$  is  $s^{\text{maj}}$ , since  $s^{\text{maj}}$  chooses the alternative that will result in the largest group of voters who get utility 1 for each realization of preferences over  $a$  and  $b$  (see the concluding remarks). Thus, if some voter's expected utility is increased by moving to an  $s$  that is higher than  $s^{\text{maj}}$ , then some other voter's expected utility must fall as the result of such a move. The same is true in reverse. So there is at least one voter with a peak at least as high as  $s^{\text{maj}}$  and at least one voter with a peak no higher than  $s^{\text{maj}}$ . The complete proof, taking into account the possibility of twin-peaks appears in the appendix.

Note that by combining Corollary 1 with Lemma 3, we know that the voter who has the highest  $p_i$  must have a  $\hat{s}_i$  which is no higher than  $s^{\text{maj}}$  and the voter who has the lowest  $p_i$  must have a  $\hat{s}_i$  that is at least as high as  $s^{\text{maj}}$ , *and this is true regardless of  $p$ .*

Finally, we also note the following lemma due to Badger (1972) and Curtis (1972), and first conjectured by Rae (1969). It singles out majority rule as a rule of special interest.

**LEMMA 4** [Badger (1972), Curtis (1972)] *For any society (profile of  $p_i$ 's), the only voting rules that maximize the sum of voters' expected utilities is  $s^{\text{maj}}$  if  $n$  is odd, and  $s^{\text{maj}}$  and  $s^{\text{maj}} - 1$  if  $n$  is even.*

While the proofs of Badger (1972) and Curtis (1972) are involved, there is a very easy way to prove Lemma 4. Given any realization of voters' preferences at time 2, the choice which maximizes the realized total utility is simply to choose the alternative preferred by a majority. Given that this is the best that one can do realization by realization, it is maximizing in total expectation as well. Any rule other than majority rule (except  $\frac{n}{2}$  when  $n$  is even) realizes a lower total utility at some realization of preferences as it will select one of the alternatives when a minority supports it, and thus we have the uniqueness claim.

## 4 Self-Stability

With some understanding of voters' preferences over voting rules under our belts, we examine the issue of existence of self-stable voting rules and constitutions in some detail.

### Existence of Self-Stable Voting Rules

We begin by considering the special case where all voters have the same  $p_i$ . This is of some interest where this common  $p$  is an indicator of the average propensity to favor change of a society's representative voter. It is also worth considering as an exercise, since the reasoning required for this simple case extends to the analysis of more heterogeneous societies. Moreover, the conclusion we reach may seem counterintuitive at first, although it is easy to reach after some reflection. The following result is a corollary of Lemmas 1 and 3.

**THEOREM 1** *If  $p_i = p_j$  for all  $i$  and  $j$ , then  $s^{\text{maj}}$  is the unique self-stable voting rule if  $n$  is odd. If  $n$  is even, then there are two self-stable rules  $s^{\text{maj}}$  and  $s^{\text{maj}} - 1$ .*

Thus, majority rule is the unique self-stable voting rule whenever all voters have the same probability of choosing change, irrespective of what this probability might be. One might have guessed that societies where all voters are very likely to want changes would prefer low values of  $s$ , that is low barriers to change, and that homogeneously conservative societies would favor high values of  $s$ . But this is not the case. Actually, in homogeneous societies, all voters have their peak at  $\hat{s}_i = s^{\text{maj}}$ , and thus majority rule is the consensus choice of rule. What actually matters is not the absolute values of the  $p$ 's but their values relative to those of other voters. For instance, consider a society where  $p_i = .01$  for each  $i$  and so voters are very conservative and very likely to support the status-quo. In this case, shouldn't it be that voters all prefer a high quota  $s$  as they each know they are likely to support the status quo? The answer is no and the reasoning lies in the answer to the following question. Which alternative would a voter prefer society to choose in a generic realization where  $k$  voters end up supporting  $a$  and  $n - k$  voters end up supporting  $b$ ? That is, the voter can think of the different scenarios possible for numbers of voters supporting  $a$  and  $b$ , and then ask which side he is most likely to fall on in each scenario. Given the symmetry in  $p_i$ 's, conditional on this realization of preferences it is most likely that the voter is in the larger of the two groups. So, the voter would like society to choose  $a$  in scenarios where  $k > n - k$  and society to choose  $b$  in scenarios where  $k < n - k$ , and is indifferent if  $k = n - k$ . Thus, the voter would like society to choose in favor of the majority as that is where the voter is most likely to be in any realization. Once one understands the above reasoning, then Lemma 3 and the importance of relative comparisons becomes clear.

Theorem 1 offers an encouraging starting point, as we find that not only does a self-stable voting rule exist, but actually the unique such rule is the efficient majority rule.

Unfortunately, the substantial symmetry in a homogeneous society is responsible for the nice conclusion of the result. In more heterogeneous societies, one can lose majority rule as being self-stable, and one can also lose existence of a self-stable voting rule altogether.

To see an example where there exist self-stable voting rules, but where majority rule is not self-stable, reconsider Example 1. Recall that the society in that example consisted of two groups of voters,  $N^1 = \{1, \dots, 4\}$  and  $N^2 = \{5, \dots, 10\}$ , where the corresponding peaks of preferences over voting rules were  $\hat{s}^1 = 8$  and  $\hat{s}^2 = 4$ . There,  $\{7, 8\}$  is the set of self-stable voting rules. It is easy to see that 8 is self-stable as only group  $N^2$  would like to change voting rules if 8 is used, but then they only have 6 members and so are too small to make the change under a rule of 8. The same is true of quota 7, and although in that case group  $N^1$  would like to raise the quota from 7 to 8 it is too small to do so. To see that no other rule is stable, note that 4 is unanimously preferred to any smaller rule, and 8 is unanimously preferred to any larger rule. So the only other candidates for self-stability are the quotas 4, 5, and 6. However, 5 and 6 are not stable because  $N^2$  prefers 4 and has enough voters to move the quota to 4. 4 is not stable since group  $N^1$  would have enough voters to increase the quota.

As we will see in Theorem 2, existence of a self-stable voting rule is guaranteed in a society where there are only two different types of voters. However, as the following example shows, existence can fail in a society with three or more types of voters.

**EXAMPLE 2** *A Society for which No Rule is Self Stable.*

$N = \{1, \dots, 5\}$ .  $p_1 = p_2 = p_3 = 1/2$ ,  $p_4 = 3/8$ , and  $p_5 = 3/16$ .

Direct calculations lead to  $\hat{s}_1 = \hat{s}_2 = \hat{s}_3 = 2$ ,  $\hat{s}_4 = 3$  and  $\hat{s}_5 = 4$ . Let us verify that there is no self-stable voting procedure. All voters want to raise the quota from 1 and lower it from 5. That leaves the quotas of 2, 3, and 4 to be checked as the only possibilities for self-stable voting rules. Voters 1 to 3 would vote to lower it from 3 to 2, voters 1 to 4 would vote to lower it from 4 to 3, and voters 3 and 4 would vote to raise it from 2 to 3. Thus, no voting rule is self-stable.

The possibility that a society may not have a self-stable voting rule is striking. In order to understand its implications, it is worth discussing more extensively when this phenomenon can or cannot occur.

There is actually much that we can deduce about the existence and properties of self-stable rules, and we collect some of this in Theorem 2. Before we state the theorem, we introduce some useful definitions.

A society  $(N, p)$  is *dichotomous* if there exists  $N^1 \neq \emptyset$ ,  $p^1 \in (0, 1)$ ,  $N^2 \neq \emptyset$ , and  $p^2 \in (0, 1)$  such that  $N = N^1 \cup N^2$ ,  $p_i = p^1$  for all  $i \in N^1$ ,  $p_i = p^2$  for all  $i \in N^2$ .

A dichotomous society is thus one that can be divided into two groups such that members of the same group have the same  $p_i$ 's, as in Example 1.

Say that a society is *symmetric* if when voters are labeled such that  $p_i \geq p_j$  when  $i > j$ , it follows that  $p_i = 1 - p_{n-i}$ .

Let  $\hat{s}_{\text{med}}$  denote the median of  $(\hat{s}_1, \dots, \hat{s}_n)$ , i.e., the median of the peaks of the voters.

**THEOREM 2** (1) *If  $\hat{s}_{\text{med}} \geq s^{\text{maj}}$ , then  $\hat{s}_{\text{med}}$  is self-stable.*

- (2) If there does not exist a self-stable voting rule for a society  $(N, p)$ , then there exists a self-stable voting rule for the society  $(N, \bar{p})$ , where  $\bar{p}$  is defined by  $\bar{p}_i = 1 - p_i$  for each  $i$ . Moreover,  $\hat{s}_{\text{med}}$  is self-stable for society  $(N, \bar{p})$ .
- (3) A dichotomous society has at least one self-stable voting rule.
- (4) If a society is symmetric, then  $s^{\text{maj}}$  is a self-stable voting rule.

Let us sketch the proof of Theorem 2 here, and we collect details in the appendix.

To see (1), note that if  $\hat{s}_{\text{med}} \geq s^{\text{maj}}$ , then at most half of the population would like to lower the rule below the median, and at most half would like to increase it above the median. Since  $\hat{s}_{\text{med}} \geq s^{\text{maj}} \geq \frac{n+1}{2}$ , it follows that  $\hat{s}_{\text{med}}$  be self-stable.

The proof of (2) follows from the observation that the setting we are examining is symmetric in the following way: if in society  $(N, p)$  voter  $i$  would like society to choose  $b$  conditional only on knowing that  $s$  voters out of society favor  $b$ , then in society  $(N, \bar{p})$  voter  $i$  would like society to choose  $a$  conditional only on knowing that  $s$  voters out of society favor  $a$ . This implies that if  $\hat{s}_i$  is  $i$ 's peak under society  $(N, p)$ , then  $n - \hat{s}_i + 1$  is  $i$ 's peak under society  $(N, \bar{p})$ . To establish (2), note that non-existence of a self-stable voting rule implies that  $\hat{s}_{\text{med}}$  is no larger than  $\frac{n}{2}$ , as otherwise it would be self-stable. The reasoning above then implies that  $\hat{s}_{\text{med}}$  for society  $(N, \bar{p})$  is larger than  $\frac{n}{2}$ , and so is stable.

(4), which asserts the existence of self-stable voting rules for symmetric societies, is an easy corollary of (2).

The proof of (3) is the most complicated of the four. It appears in the appendix and involves explicit examination of voters' conditional probabilities that they will support alternative  $b$  if  $k$  voters support  $b$ . Very roughly, it works by relating the conditional beliefs of the two groups to each other. Let  $N^1, N^2, n^1, n^2, \hat{s}^1$ , and  $\hat{s}^2$  be the two groups of voters, the cardinalities of these groups, and their peaks, respectively. The main case that has to be ruled out to establish existence is where  $n^2 \geq \hat{s}^1$  and  $n^1 \geq \hat{s}^2$ , when  $\hat{s}^1 \neq \hat{s}^2$ .<sup>11</sup> If the beliefs of  $N^1$  are such that  $n^2 \geq \hat{s}^1$ , this means that the voters in  $N^1$  have *relatively* high beliefs that they will be among the supporters of  $b$ . This implies that the voters in  $N^2$  have *relatively* low beliefs that they will be among the supporters of  $b$ , and so  $\hat{s}^2$  will be high enough to be larger than  $n^1$ . The challenge in the proof is to show that these relative statements translate into absolute statements about the relationship between  $\hat{s}^1$  and  $\hat{s}^2$  and their comparison to  $n^1$  and  $n^2$ .

---

<sup>11</sup>If, for instance,  $\hat{s}^1 > n^2$  then  $\hat{s}^1$  would be self-stable. So it would have to be that both  $n^2 \geq \hat{s}^1$  and  $n^1 \geq \hat{s}^2$  for there not to exist a self-stable rule. Without loss of generality let  $\hat{s}^2 > \hat{s}^1$ , as the case where  $\hat{s}^2 = \hat{s}^1$  would lead to unanimity and thus self-stability. So, to see that if a case existed where  $n^2 \geq \hat{s}^1$  and  $n^1 \geq \hat{s}^2$  when  $\hat{s}^1 \neq \hat{s}^2$ , then there would not exist a self-stable voting rule, note that there would be unanimous support for change of any  $s$  that lies outside of the range between (and including)  $\hat{s}^1$  and  $\hat{s}^2$ . Also  $N^1$  would want to change away (and could change) from and  $s$  such that  $\hat{s}^2 \geq s > \hat{s}^1$ . Finally,  $N^2$  would want to change from  $\hat{s}^1$ .

(2) has some powerful implications. It implies that non-existence is a problem for less than “half” of the potential societies, in terms of the  $p$ 's. This implies that while non-existence can occur for open sets of societies (simply build a neighborhood around Example 2), it still is a problem that is not completely pervasive.

We also note that when self-stable voting rules exist, there may be a number of them. Moreover, the set of self-stable voting rules need not be an interval, nor need it include  $s^{\text{maj}}$ . These points are illustrated in the following example.

**EXAMPLE 3** *A Society with Multiple and Non-Adjacent Self Stable Rules.*

The society  $(N, p)$  is dichotomous.

$N^1 = \{1, \dots, 5\}$  and  $N^2 = \{6, \dots, 16\}$  with  $p^1 = .01$  and  $p^2 = .99$ .

Here  $\hat{s}^1 = 14$  and  $\hat{s}^2 = 6$ .

It follows that  $\{6, 12, 13, 14\}$  is the set of self-stable voting rules.

It is clear that the set of self-stable voting rules will consist of a set of intervals, each of which includes at least one  $\hat{s}_i$ . This puts an upper bound on the number of disjoint intervals that can be included, at the number of distinct  $p_i$ 's that are present in the society.

### Sub-Majority Voting Rules

Before turning to the question of constitutional design, let us comment on some problems related to the choice of majority sizes smaller than  $s^{\text{maj}}$ .

Rules with  $s < s^{\text{maj}}$  can be problematic in the following sense. Consider a situation where  $a$  and  $b$  are each supported by half of the population. A vote under  $s$  will result in  $b$  becoming the new status quo. But then, with  $b$  as the new status quo, the other half of the voters would support (and could effect) change back to  $a$  if it is proposed for a vote against  $b$ . Thus, there is the potential to continuously cycle back and forth between  $a$  and  $b$  as the status quo.<sup>12</sup> This, of course, is only a potential problem of sub-majority rules.<sup>13</sup>

Suppose that a society somehow precludes itself from ever selecting a sub-majority rule. If this is the case, then the existence of self-stable voting rules is ensured. To see this, consider such a society. The preferences of voters over the restricted set of  $s$ 's ( $s \geq s^{\text{maj}}$ ) are still single peaked. Voters whose unrestricted peaks were at least  $s^{\text{maj}}$  have the same peak

<sup>12</sup>See Caplin and Nalebuff (1988) for a related discussion of the instability of infra-majority rules.

<sup>13</sup>Note that there are two caveats to the above noted difficulty with sub-majority rules. First, for some alternatives it may not be possible to make reversals. For instance, if  $a$  is a current membership of a society and  $b$  is a question to include a new member, it may not be permitted to later vote to revoke membership. There are many such examples of decisions which cannot be reversed, such as a vote to tenure a faculty member, or a vote to declare war, etc.. Second, the difficulty requires that one reasonably expect that the reversed proposal be made, and so the agenda control becomes important. It may be that the agenda is controlled in manners so that once  $b$  has been voted for,  $a$  is never again pitted as an alternative. We have abstracted away from the agenda in our model, and a more complete analysis of the potential instability of sub-majority rules demands a careful modeling of the agenda.

on the restricted set, while voters whose peaks were below  $s^{\text{maj}}$  now have  $s^{\text{maj}}$  as a peak. The median of the restricted peaks will be self-stable over the restricted set of voting rules. This leads to the following theorem.

**THEOREM 3** *For any society where only  $s \geq s^{\text{maj}}$  are admissible voting rules,  $\hat{s}_{\text{med}}$  (defined relative to restricted preferences) is a self-stable voting rule.*

Theorem 3 follows as a corollary of (1) of Theorem 2. Hence, for societies who exclude  $s$ 's below  $s^{\text{maj}}$  a priori, we can add self-stability to the list of properties in the literature justifying median voting rules.

It is interesting to note that even when sub-majority rules are ruled out, it can be that the efficient rule, majority rule, is not self-stable. Actually, from (2) in Theorem 2, we can deduce that majority rule will turn out to be self-stable in at least “half” of the societies, in terms of the possible  $p$ 's. And we also know that  $s^{\text{maj}}$  will not be self-stable when  $\hat{s}_{\text{med}} > s^{\text{maj}}$ .

Finally, notice that in Example 3, a rule with  $s$  below  $s^{\text{maj}}$  ( $s=6$ ) emerges as self-stable, along with others involving values above majority.<sup>14</sup> Excluding these low value rules a priori will deprive us of knowing all possible stable arrangements, when they exist.

### Self-Stable Constitutions

As we have seen so far, self-stable voting rules will exist for many, but not all, societies. Does this mean we should take the possibility that a society might not find a self-stable rule as a serious threat to the stability of decision making? One answer is that this possibility of instability helps explain why many (if not most) societies resort to special rules when it comes to changing the voting rules. This motivates an analysis of self-stable constitutions.

Existence is now guaranteed.

**THEOREM 4** *For any society, the constitutions  $(s^{\text{maj}}, n)$  and  $(\hat{s}_{\text{med}}, S)$  for any  $S \geq s^{\text{maj}}$  are self-stable.*

Theorem 4 follows is a straightforward consequence of our results on intermediate preferences (Lemma 2) and on relative positioning of voter's peaks (Lemma 3), and so we simply offer a description of the proof as follows. The self-stability of  $(s^{\text{maj}}, n)$  follows from the observation that by Lemmas 3 and 2 there is always at least one voter who will wish to keep the voting rule over issues no higher than  $s^{\text{maj}}$  and at least one who will wish to keep the voting rule no lower than  $s^{\text{maj}}$ . Thus, there is no unanimous consent to raise or lower the voting rule from  $s^{\text{maj}}$ . The self-stability of  $(\hat{s}_{\text{med}}, S)$  with  $S \geq s^{\text{maj}}$  follows from Lemma 2 and the definition of  $\hat{s}_{\text{med}}$ , as by intermediate preferences fewer than  $n/2$  voters will prefer

---

<sup>14</sup>Moreover, there exist examples where the only self-stable rules are sub-majority rules. For example, consider a dichotomous society with  $N^1 = \{1, 2\}$  and  $N^2 = \{3, \dots, 7\}$  with  $p^1 = .3$  and  $p^2 = .5$ . Straightforward calculations lead to  $\hat{s}^1 = 5$  and  $\hat{s}^2 = 3$ . There  $s = 3$  is the only self-stable rule.



to raise the voting rule from  $\widehat{s}_{\text{med}}$ , and similarly fewer than  $n/2$  voters will prefer to lower the voting rule from  $\widehat{s}_{\text{med}}$ .

Theorem 4 is essentially tight in the sense that for any  $(s, S)$  that does not coincide with either  $(s^{\text{maj}}, n)$  or  $(\widehat{s}_{\text{med}}, S)$  with  $S \geq s^{\text{maj}}$ , there is some situations in which  $(s, S)$  is not self-stable (with a single exception that  $(s^{\text{maj}}, n - 1)$  is always self-stable whenever  $n$  is odd).<sup>15</sup> Let us be more explicit. First, consider  $(s, S)$  with some  $s \neq s^{\text{maj}}$ . If  $p$  is such that  $p_i = p_j$  for all  $i$  and  $j$  then  $\widehat{s}_i = s^{\text{maj}}$  for all  $i$  and so any  $(s, S)$  for which  $s \neq s^{\text{maj}} = \widehat{s}_{\text{med}}$  will be unstable regardless of  $S$ . So we need only consider  $(s, S)$  where  $s = s^{\text{maj}}$  or  $s = \widehat{s}_{\text{med}}$ . We can see the problem with  $(\widehat{s}_{\text{med}}, S)$  where  $S < s^{\text{maj}}$  from Example 3, as it is possible to have societies where a near majority prefers to move the voting rule away from  $\widehat{s}_{\text{med}}$ .<sup>16</sup> Finally, when considering  $(s^{\text{maj}}, S)$  with  $S < n$  ( $S < n - 1$  if  $n$  is odd), consider a society where voter 1 has  $p_1$  near 0 (and the same for voter 2 in the case of  $n$  being odd), and all other voters have the same  $p_i$  near 1. For high enough  $p_i$ , voters  $i$  will have probability greater than  $1/2$  of supporting change when there are  $n/2$  supporters of change if  $n$  is even and when there are  $(n - 1)/2$  supporters of change when  $n$  is odd. This leads to peaks of  $s^{\text{maj}} - 1$  for the voters with  $p_i$  near 1, and so they will vote to decrease the voting rule if it is set at  $s^{\text{maj}}$ .

In summary. For each society, there will always be at least two self-stable constitutions (three when  $n$  is odd).

Although we have treated the constitutions  $(s^{\text{maj}}, n)$  and  $(\widehat{s}_{\text{med}}, S)$  on equal footing in the statement of Theorem 4, notice the following essential difference. The constitution  $(\widehat{s}_{\text{med}}, S)$  varies across societies, since  $\widehat{s}_{\text{med}}$  depends on the distribution of  $p_i$ 's. On the other hand,  $(s^{\text{maj}}, n)$  is the same across all societies of the same size. Hence,  $(s^{\text{maj}}, n)$  is a stable constitution regardless of the society, while a constitution of the form  $(\widehat{s}_{\text{med}}, S)$  is by definition tailored to a specific society.

## Majority Rule Constitutions

The self-stability of constitutions using majority rule as a voting is of particular interest because of the prominence of majority rule in actual constitutions and its special properties including overall efficiency (Theorem 4). We have just seen that the particular constitution  $(s^{\text{maj}}, n)$  is self-stable for any society.

We now explore the conditions on the distribution of  $p_i$ 's that are sufficient for other constitutions  $(s^{\text{maj}}, S)$  to be self-stable for values of  $S < n$ .

Let us say a couple of words about why we should care about these constitutions. Here, we have not modeled why a society should every want to set  $S < n$ . Nevertheless, societies do desire flexibility and we see that most constitutions allow for the real possibility of amendment. Regardless of their motivation, we are still interested in understanding which

---

<sup>15</sup>Note that any Pareto optimal  $s$  is stable when put together with  $n$ . The claim here is that  $s^{\text{maj}}$  is the only  $s$  that is Pareto optimal for all societies.

<sup>16</sup>More generally, consider a society with a single voter who has the median preferences and other voters who have extreme  $p_i$ 's near 0 and 1, who will prefer to lower or raise the voting rule. In particular, the voters with  $p_i$ 's near enough to 1 will prefer an  $s < \widehat{s}_{\text{med}}$  over  $\widehat{s}_{\text{med}}$  and there will be at least  $s^{\text{maj}} - 1$  such voters.

constitutions are self-stable, as those are the ones that we should expect to last for “long” periods of time; where “long” is relative to the length of time that a society has some continuity in its views.

It is important to note that Example 3 provides an example where a constitution of  $(s^{\text{maj}}, \frac{2}{3}n)$  is not self-stable. There, 11 of the 16 agents would prefer to lower the voting role to be below  $s^{\text{maj}}$ . Thus, there are simple situations where seemingly natural constitutions are not self-stable. The following theorem provides a characterization of which constitutions involving majority rule are self-stable, as a function of the society.

Let  $z_i = \frac{p_i}{1-p_i}$ . Thus,  $z_i$  represents the ratio of the probability that  $i$  supports change compared to the probability that  $i$  supports the status quo. Any positive number is a potential  $z_i$ .

**THEOREM 5** *For any society with even  $n$  the constitution  $(s^{\text{maj}}, S)$  is self-stable if*

$$S > |\{i : \sum_{C \subset N, |C|=\frac{n}{2}, i \in C} (\times_{j \in C} z_j) \geq \sum_{C \subset N, |C|=\frac{n}{2}, i \notin C} (\times_{j \in C} z_j)\}| > n - S. \quad (3)$$

Note that (3) can be rewritten as

$$S > |\{i : z_i \geq \sum_{k \neq i} \lambda_k^i z_k\}| > n - S, \quad (4)$$

where

$$\lambda_k^i = \frac{2 \sum_{|C|=\frac{n}{2}-1, i, k \notin C} (\times_{j \in C} z_j)}{n \sum_{|C|=\frac{n}{2}-1, i \notin C} (\times_{j \in C} z_j)}.$$

Here, the  $\lambda_k^i$  are weights such that  $\sum_{k \neq i} \lambda_k^i = 1$ , and so  $\sum_{k \neq i} \lambda_k^i z_k$  is a weighted average of  $z_k$ 's over  $k$ 's other than  $i$ . Thus, (4) says roughly that the number of voters with above average  $z_i$ 's is not too high and not too low. It can be shown that this is also equivalent to having the number of voters with below average  $z_i$ 's not be too high or too low.

Condition (3) is almost a necessary condition as well, except for the possibility that one particular voter (the  $n - S$ -th voter when ordered in terms of decreasing  $p_i$ 's) has peak exactly at  $s^{\text{maj}}$  which allows for a slightly weaker condition.

To see the implications of Theorem 5, let us consider the constitution where  $s' = 2n/3$ . That constitution is stable, provided there are at least  $1/3$  of the voters who do not wish to raise the voting rule from  $s^{\text{maj}}$  and at least  $1/3$  of the voters who do not wish to lower it from  $s^{\text{maj}}$ . The proof of the theorem involves showing that these are equivalent to the inequalities relating the  $z_i$ 's. The requirements of the theorem are then that at least  $1/3$  and no more than  $2/3$  of the voters have a  $z_i$  that is bigger than the weighted average of the other voters'  $z_i$ 's. This is in effect a limitation on the skewness of the distribution of the  $z_i$ 's (or, in effect, the  $p_i$ 's). If the distribution of  $z_i$ 's is not too skewed, then  $(s^{\text{maj}}, 2n/3)$  will be self-stable.

More generally, Theorem 5 provides the reasoning behind why a super-majority will be required for rules changes in a constitution where majority rule is used for ordinary decisions.

## 5 Concluding Remarks

Our research takes the view that ‘choosing how to choose’ is an issue that calls for the treatment of institutions as endogenous variables at equilibrium, and not as exogenously given data. Thus, we see it as part of a broad and ambitious research program of not only understanding normative or positive properties of institutions and mechanisms, but also how they come to take certain forms when individuals in the society have personal stakes in the design of the institution and can affect it. To some extent this presents a ‘chicken and egg’ dilemma, as the existing institutional environment to a large extent determines what institutional changes can take place, and are also the result of previous institutional change. Economics has a tradition of dealing with problems of this kind by resorting to appropriate fixed-point and equilibrium notions, and self-stability can be viewed in this light.

In order to study self-stability, we have considered a model that we realize is stylized on many dimensions, such as taking the agenda to be a binary one, taking the agenda to be exogenous, examining only anonymous voting rules, considering non-repeated environments, and considering a fixed population of voters. Nevertheless, we feel that it provides important steps in modeling the very important process of how a society chooses the institutions that it uses for governance. Relaxing some of these restrictions provides a rich agenda for further research.

### **Anonymity**

Anonymity has been presumed in our analysis through the definition of a voting rule: every voter has an equal weight. But in many cases voters are not individuals with equal rights. Rather, they may be countries of different sizes (as in the council of the European Union), or government agencies who contribute differently to a shared institution (as in the IMF). In many such cases, different voters get different weights. Again, these weights may eventually be changed through a vote. The analysis of self-stable constitutions for rules of this type is important and certainly nontrivial, as shown by some initial results of Sosnowska (2002). In a recent paper (Barbera and Jackson (2003)) we study alternative rationales for the choice of weights, and their connection with the population variable.

### **Different Rules for Different Issues**

Here we have analyzed a choice of a single voting rule. Generally, the different types of issues a society faces might have different characteristics (for instance, different voter  $p_i$ 's), and that might lead to different choices of voting rules for different issues. For instance, in California propositions involving new bond issues are often held to higher majorities than other propositions.<sup>17</sup>

### **Large Numbers**

---

<sup>17</sup>See Tsebelis and Money (1997) for some interesting descriptions of variations on rules for different sorts of decisions.

We have deliberately worked with finite societies for two reasons. First, there are many applications where the society in question is small and not well approximated by an infinite society. Second, if one worked with a continuum society (or some other infinite model), then, without making additional assumptions about the distribution of the underlying uncertainty, a (suitable) law of large numbers would eliminate the uncertainty over the proportion of society supporting change over the status quo. This uncertainty is the critical aspect that makes for non-degenerate and interesting voters' preferences over voting rules.

While we chose to work with a finite model, it is still interesting to ask questions about large societies.<sup>18</sup> Example 3 extends when the society is replicated a large number of times, and so general existence of self-stable voting rules will not come simply from considering a large society. However, there may be some interesting conditions that are sufficient for self-stability one can obtain from looking at large societies.<sup>19</sup>

## Preferences

Assuming that a voter gets a utility of 1 when his preferred alternative is selected and 0 otherwise involves more than a normalization. Instead, it could be that when voter  $i$  supports  $a$  then  $i$  gets utility 1 if  $a$  is selected and 0 if  $b$  is selected, while when voter  $i$  supports  $b$  then  $i$  gets utility  $x_i$  when  $b$  is selected and 0 if  $a$  is selected.

This more general setting leads to changes in the analysis in the following ways.

First, Lemma 1 on single-peaked preferences goes through unaltered and it is easily checked that the proof works with only slight modification.

Second, the extension of Lemma 2 on intermediate preferences is a more complicated matter. There are now two characteristics that distinguish voters and so finding an ordering on voters for which their preferences are intermediate is more delicate. In the case where  $x_i \geq x_j$  whenever  $p_i \geq p_j$  (so that voters who are more likely to support alternative  $b$  care relatively more about alternative  $b$ ), preferences are still intermediate. Again, for this situation the proof goes through with very little modification. This would seem to be a natural condition. However, if there is no such relationship between the  $x_i$ 's and the  $p_i$ 's, then preferences may fail to be intermediate, and it is easy to construct counter-examples.

The existence of self-stable voting rules with dichotomous preferences, Theorem 2 (3), depends on the property that a voter cares (in expectation, at least) similarly for having  $a$  win when the voter supports  $a$  and having  $b$  win when the voter supports  $b$ . Without that assumption, examples can be constructed where there does not exist a self-stable voting rule. However, Theorem 2 (4) extends under an ordering that preserves the intermediate preferences.

Another aspect of preferences that might be due for further consideration is the assumption of the independence of the probabilities that the voters support change. This assumption played a role in our proof of single-peakedness. Most importantly, this ensures that likelihood

---

<sup>18</sup>See Schofield (1971) for some calculations concerning voters' preferences in large heterogeneous societies.

<sup>19</sup>One possibility is to think about conditions on the distributions of  $p_i$ 's, in an analogous way that conditions identified by Caplin and Nalebuff (1988) on distributions of preferences suffice for an alternative with nice properties in their setting.

that a voter has that they support change conditional on  $k$  voters supporting change is be monotone in  $k$ . With certain forms of correlation, this conditional probability may no longer be monotone. While arbitrary forms of correlation could be difficult to accommodate, there are natural ones which still allow for such monotonicity and would thus still be tractable.

## Empirical Investigations

As has been pointed out to us, one can test some of the ideas developed here rather directly as follows. One could look at roll-call data from the U.S. Senate to see who voted for proposed rules changes. One could then compare this to the votes by senators on previous bills (which would provide an estimate of the various  $p_i$ 's). Senators with relatively higher propensities to support change, should also be opposing proposals (which have appeared from time to time) to raise the rule for ending a debate and calling a vote from 6/10 to 2/3, as this is effectively an increase in  $s$ .

Of course, this is just one possible empirical investigation, and one that provides more of a specific check on some of the predictions of the model. More generally, developing an understanding of how the stability of a constitution relates to the underlying primitives of the society provides a rich agenda for investigation.

## References

- Aghion, P., A. Alesina, and F. Trebbi (2002) "Endogenous Political Institutions," mimeo: Harvard.
- Austen-Smith, D. and J. Banks (1996) "Information Aggregation, Rationality and the Condorcet Jury Theorem," *American Political Science Review*, vol. 90, pp. 34–45.
- Badger, W.W. (1972) "Political Individualism, Positional Preferences, and Optimal Decision-Rules," in *Probability Models of Collective Decision Making*, edited by R.G. Niemi and H.F. Weisberg, Merrill Publishing: Columbus Ohio.
- Barbera, S. and M.O. Jackson (2003) "On the Weights of Nations: Assigning Voting Weights in a Heterogeneous Union," <http://www.hss.caltech.edu/~jacksonm/Jackson.html>
- Buchanan, J.M. and G. Tullock (1962) *The Calculus of Consent: Logical Foundations of Constitutional Democracy* University of Michigan Press: Ann Arbor.
- Caplin, A. and B. Nalebuff (1988) "On 64%-Majority Rule." *Econometrica*, vol. 56, pp 787–814.
- Coelho, D. (2002) "Maximin Choice of Voting Rules for Committees," mimeo: Universitat Autònoma de Barcelona.
- Condorcet, Marquis de (1785) *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris.

- Curtis, R.B. (1972) "Decision Rules and Collective Values in Constitutional Choice," in *Probability Models of Collective Decision Making*, edited by R.G. Niemi and H.F. Weisberg, Merrill Publishing: Columbus Ohio.
- Dasgupta, P. and E. Maskin (1998) "Robustness of Majority Rule," mimeo.
- Feddersen, T.J. and W. Pesendorfer (1998) "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts," *American Political Science Review*, vol. 92, pp. 23–36.
- Grandmont, J.M. (1978) "Intermediate Preferences and the Majority Rule," *Econometrica*, vol. 46, pp. 317–330.
- Koray, S. (2000) "Self-Selective Social Choice Functions Verify Arrow and Gibbard-Satterthwaite Theorems," *Econometrica*, vol. 68, pp. 981–996.
- Niemi and Weisberg (1972) "Substantive Applications of Collective Decision-Making," in *Probability Models of Collective Decision Making*, edited by R.G. Niemi and H.F. Weisberg, Merrill Publishing: Columbus Ohio.
- Persson, T. (2002) "Do Political Institutions Shape Economic Policy?" *Econometrica*, 70, pp. 883-906.
- Persson, T. and G. Tabellini (2000) *Political Economics: Explaining Economic Policy*, Cambridge: MIT Press.
- Polborn, M. and M. Messner (2002) "Voting on Majority Rules," mimeo: Pompeu Fabra.
- Rae, D. (1969) "Decision Rules and Individual Values in Constitutional Choice," *American Political Science Review*, vol. 63, pp. 40–56.
- Rousseau, J.-J. (1762) *du Contrat Social; ou, Principes du Droit Politique* reprinted: Paris : Garnier; 1966.
- Schofield, N.J. (1971) "Ethical Decision Rules for Uncertain Voters," *British Journal of Political Science*, vol. 2, pp 193–207.
- Schofield, N.J. (1972) "Is Majority Rule Special?" in *Probability Models of Collective Decision Making*, edited by R.G. Niemi and H.F. Weisberg, Merrill Publishing: Columbus Ohio.
- Sosnowska, Honorata (2002) "Self-Stability for Weighted Voting Games," mimeo: Warsaw School of Economics.
- Tsebelis, G. and J. Money (1997) *Bicameralism*, Cambridge University Press: Cambridge.

Wakayama, T. (2002) “Endogenous Choice of Voting Rules with Abstention,” mimeo: Osaka University.

Young, H.P. and A. Levenglick (1978) “A Consistent Extension of Condorcet’s Election Principle” *SIAM Journal of Applied Mathematics, Part C*, vol. 35, pp. 283–300.

## Appendix

### Proof of Lemma 1:

Let  $D_i(s) = U_i(s) - U_i(s - 1)$ . From equation (2) it follows that

$$D_i(s) = (1 - p_i)P_i(s - 1) - p_iP_i(s - 2). \quad (5)$$

Thus,

$$D_i(s) = P_i(s - 1) \left( 1 - p_i \frac{P_i(s - 1) + P_i(s - 2)}{P_i(s - 1)} \right).$$

Note that  $U_i$  is single-peaked if there exists  $\hat{s}_i$  (possibly equal to 1 or  $n$ ) such that  $D_i(s) > 0$  for every  $\hat{s}_i > s \geq 2$ ,  $D_i(s) < 0$  for every  $n \geq s > \hat{s}_i$ , and  $D_i(s) \geq 0$  at  $s = \hat{s}_i$  (with equality holding only when there are twin peaks). Thus, if we can show that  $D_i(s)$  has this form, then we will have shown that  $U_i$  is single-peaked.

Note that the sign of  $D_i(s)$  depends only on the size of  $\frac{P_i(s-1)+P_i(s-2)}{P_i(s-1)}$  relative to  $\frac{1}{p_i}$ . This means that showing that  $\frac{P_i(s-1)+P_i(s-2)}{P_i(s-1)}$  is increasing in  $s$  for  $n \geq s \geq 2$  establishes that  $D_i(s)$  has the form specified above. Rewriting

$$\frac{P_i(s - 1) + P_i(s - 2)}{P_i(s - 1)} = 1 + \frac{P_i(s - 2)}{P_i(s - 1)},$$

means that we need only show that  $\frac{P_i(s-2)}{P_i(s-1)}$  is increasing<sup>20</sup> in  $s$ .

We follow a proof by induction on  $n$ . The case where  $n = 2$  is trivial, since then there is only one  $s$  that satisfies  $n \geq s \geq 2$ . Now for the induction step. Suppose that  $P_i(s - 2)/P_i(s - 1)$  is increasing for any  $n' \geq s \geq 2$  for societies of size  $n - 1 \geq n'$ . We show that  $P_i(s - 2)/P_i(s - 1)$  is increasing for any  $n \geq s \geq 2$ .

Let

$$P_{i,j}(s) = \sum_{C \subset N \setminus \{i,j\}: |C|=s} \times_{\ell \in C} p_\ell \times_{k \notin C} (1 - p_k).$$

$P_{i,j}(s)$  is the probability that exactly  $s$  of the voters other than  $i$  and  $j$  support the change.

$$\frac{P_i(s - 2)}{P_i(s - 1)} = \frac{p_j P_{i,j}(s - 3) + (1 - p_j) P_{i,j}(s - 2)}{p_j P_{i,j}(s - 2) + (1 - p_j) P_{i,j}(s - 1)},$$

---

<sup>20</sup>When we say “increasing” we refer to the strict sense, and we use the term “non-decreasing” to refer to the weaker sense.

where  $P_{i,j}(s-3) = 0$  when  $s = 2$ . Rewrite the above equality as

$$\frac{P_i(s-2)}{P_i(s-1)} = \frac{p_j P_{i,j}(s-3)}{p_j P_{i,j}(s-2) + (1-p_j) P_{i,j}(s-1)} + \frac{(1-p_j) P_{i,j}(s-2)}{p_j P_{i,j}(s-2) + (1-p_j) P_{i,j}(s-1)}. \quad (6)$$

We show that each term on the right hand side of (6) is increasing in  $s$  for  $n \geq s \geq 2$ . Take the first term. It is clear that since  $P_{i,j}(s-3) = 0$  when  $s = 2$ , that it is increasing from  $s = 2$  to  $s = 3$ . So, we need only show that its inverse is decreasing in  $s$  for  $n \geq s \geq 3$ .

$$\begin{aligned} \frac{p_j P_{i,j}(s-2) + (1-p_j) P_{i,j}(s-1)}{p_j P_{i,j}(s-3)} &= \frac{p_j P_{i,j}(s-2)}{p_j P_{i,j}(s-3)} + \frac{(1-p_j) P_{i,j}(s-1)}{p_j P_{i,j}(s-3)} \\ &= \frac{p_j P_{i,j}(s-2)}{p_j P_{i,j}(s-3)} + \frac{(1-p_j)}{p_j} \frac{P_{i,j}(s-1) P_{i,j}(s-2)}{P_{i,j}(s-2) P_{i,j}(s-3)}. \end{aligned} \quad (7)$$

Note that the induction step implies that  $\frac{P_{i,j}(s-2)}{P_{i,j}(s-3)}$  is increasing in  $s$  for  $n \geq s \geq 3$ . So, each expression on the right hand side is decreasing in  $s$  for each  $n-1 \geq s \geq 3$  by the induction step, and so the overall expression is. So we only have to worry about the case where  $s = n$  and the expression  $\frac{P_{i,j}(s-1)}{P_{i,j}(s-2)}$ . Note that  $P_{i,j}(n-1) = 0$ , and so this follows as well.

Recall that the expression in (7) is the inverse of the first term on the right hand side of (6). A similar argument establishes that the second term on the right hand side of (6) is increasing in  $s$ . ■

### Proof of Lemma 2:

We first show that  $U_i(s) \geq U_i(s-1)$  implies that  $U_j(s) \geq U_j(s-1)$  for any  $j$  such that  $p_j \geq p_i$ . Recall that

$$U_i(s) - U_i(s-1) = (1-p_i)P_i(s-1) - p_i P_i(s-2).$$

So, we write

$$U_i(s) - U_i(s-1) = (1-p_i)(P_{i,j}(s-2)p_j + P_{i,j}(s-1)(1-p_j)) - p_i(P_{i,j}(s-3)p_j + P_{i,j}(s-2)(1-p_j)).$$

Likewise,

$$U_j(s) - U_j(s-1) = (1-p_j)(P_{i,j}(s-2)p_i + P_{i,j}(s-1)(1-p_i)) - p_j(P_{i,j}(s-3)p_i + P_{i,j}(s-2)(1-p_i)).$$

It follows that

$$[U_i(s) - U_i(s-1)] - [U_j(s) - U_j(s-1)] = 2(p_j - p_i)P_{i,j}(s-2).$$

Notice that the right hand side of the above equation is nonnegative, because  $p_j \geq p_i$ .

So, we have shown that

$$U_i(s) - U_i(s-1) \geq U_j(s) - U_j(s-1), \quad (8)$$



whenever  $p_j \geq p_i$ . Note that if  $s > s'$ , then

$$U_i(s) - U_i(s') = \sum_{k=s'+1}^s U_i(k) - U_i(k-1). \quad (9)$$

So, (9) implies that

$$U_i(s) - U_i(s') \geq U_j(s) - U_j(s'), \quad (10)$$

whenever  $p_j \geq p_i$ , provided  $s > s'$ . This establishes that preferences satisfy the single-crossing property.

We now show that this implies that preferences are intermediate.

First, consider the case where  $U_i(s) \geq U_i(s')$ ,  $U_j(s) \geq U_j(s')$ ,  $p_j \geq p_k \geq p_i$  and  $s > s'$ . Since  $U_i(s) \geq U_i(s')$  and  $p_k \geq p_i$ , (10) implies that  $U_k(s) \geq U_k(s')$  (with strict inequality if the strict inequality holds for  $i$ ). So, the desired conclusion of intermediate preferences is established for this case (and the corresponding strict inequality case). Next, consider the case where  $U_i(s) \geq U_i(s')$ ,  $U_j(s) \geq U_j(s')$ ,  $p_j \geq p_k \geq p_i$  and  $s < s'$ . Suppose to the contrary that  $U_k(s') > U_k(s)$ . Then since  $p_j \geq p_k$  it follows from (10) (applied with the roles of  $s$  and  $s'$  reversed) that  $U_j(s') > U_j(s)$ , which is a contradiction. Thus, our supposition was incorrect and so  $U_k(s) \geq U_k(s')$ . Finally, consider the case where  $U_i(s) > U_i(s')$ ,  $U_j(s) > U_j(s')$ ,  $p_j \geq p_k \geq p_i$  and  $s < s'$ . Suppose to the contrary that  $U_k(s') \geq U_k(s)$ . Then since  $p_j \geq p_k$  it follows from (10) (applied with the roles of  $s$  and  $s'$  reversed) that  $U_j(s') \geq U_j(s)$ , which is a contradiction. Thus, our supposition was incorrect and so  $U_k(s) > U_k(s')$ . We have shown that preferences are intermediate. ■

**Proof of Lemma 3:**  $s^{\text{maj}}$  maximizes total societal welfare (Theorem 4). Consider the case where some voter  $j$ 's peak is greater than  $s^{\text{maj}}$ . So,  $U_j(s^{\text{maj}} + 1) \geq U_j(s^{\text{maj}})$ . As  $\sum_i U_i(s^{\text{maj}}) > \sum_i U_i(s^{\text{maj}} + 1)$ , it follows that there exists some  $i$  with  $U_i(s^{\text{maj}} + 1) < U_i(s^{\text{maj}})$  which by single-peaked preferences (Lemma 1) implies that  $s^{\text{maj}} \geq \hat{s}_i$ . We are left with the case where all voters' peaks are no more than  $s^{\text{maj}}$ . Suppose to the contrary of Lemma 3 that all the peaks are strictly less than  $s^{\text{maj}}$ . By the single-peakedness of preferences, this implies that  $U_i(s^{\text{maj}} - 1) > U_i(s^{\text{maj}})$  for all  $i$ , which contradicts the fact that  $s^{\text{maj}}$  maximizes  $\sum_i U_i(s)$ . Thus our supposition was incorrect and the Lemma is established. ■

**Proof of Theorem 2:** The proofs for (1), (2), and (4) appear in the text. Let us prove (3). Let  $n^2 = \#N^2$  and  $n^1 = \#N^1$ , and without loss of generality take  $p_1 \geq p_2$ .

In the case where  $p^1 = p^2$  it is easily checked that all preferences are identical with  $\hat{s}_i = n/2 + 1$  if  $n$  is even, and  $\hat{s}_i = (n + 1)/2$  if  $n$  is odd. In that case,  $\hat{s}_i$  is self-stable. So, we consider the case where  $p^1 > p^2$ ,  $n_1 \geq 1$ , and  $n_2 \geq 1$ .

Lemma 3 and Corollary 1 implies that  $\hat{s}_2 \geq s^{\text{maj}} \geq \hat{s}_1$ , since  $p_2 < p_1$ . If  $n_2 \geq s^{\text{maj}}$ , then it must be that  $s^{\text{maj}} > n_1$  and so  $\hat{s}_2$  is self-stable. Therefore, we need only examine the case where  $n_1 \geq s^{\text{maj}} > n_2$ .

Suppose to the contrary that there is no self-stable voting rule. It must be that  $n_2 \geq \hat{s}_1$  and  $n_1 \geq \hat{s}_2$ . Thus,

$$n_1 \geq \hat{s}_2 \geq s^{\text{maj}} > n_2 \geq \hat{s}_1 \quad (11)$$

and  $p_2 < p_1$ .

For  $k \in \{1, \dots, n\}$ , let  $q_i^b(k)$  be the probability that a voter of type  $i \in \{1, 2\}$  supports  $b$  conditional on knowing that  $k$  voters support  $b$ . Correspondingly, let  $q_i^a(k)$  be the probability that a voter of type  $i \in \{1, 2\}$  supports  $a$  conditional on knowing that  $k$  voters support  $a$ . By the definition of  $q_i^a$  and  $q_i^b$  it follows that

$$q_i^a(k) = 1 - q_i^b(n - k). \quad (12)$$

Note that  $i$ 's peak is the largest  $s'$  such that  $q_i^b(s') \geq 1/2$  and  $1/2 \geq q_i^b(s)$  for  $s < s'$ .

Below we will establish that

$$\frac{q_2^a(k)}{k} > \frac{q_2^a(k+1)}{(k+1)}. \quad (13)$$

Before proving (13), let us argue that this will complete the proof. Since  $q_2^b(\hat{s}_2) \geq 1/2$  it follows that  $1/2 \geq q_2^a(n - \hat{s}_2)$ . So, by (13) (applied iteratively) it follows that

$$1/2 > q_2^a(n - \hat{s}_1) \frac{n - \hat{s}_2}{n - \hat{s}_1}.$$

From the inequality above, we then have

$$n^2 q_2^a(n - \hat{s}_1) < \frac{n_2}{2} \frac{n - \hat{s}_1}{n - \hat{s}_2}.$$

Since it must be that  $n_1 q_1^a(n - s) + n_2 q_2^a(n - s) = n - s$ , it follows that

$$n_1 q_1^a(n - \hat{s}_1) > n - \hat{s}_1 - \frac{n_2}{2} \frac{n - \hat{s}_1}{n - \hat{s}_2}.$$

Noting that  $n - \hat{s}_1 \geq n_1$  (recall that  $n_1 + n_2 = n$  and  $n_2 \geq \hat{s}_1$  from inequality (11)), the previous inequality requires that

$$q_1^a(n - \hat{s}_1) > 1 - \frac{n_2}{2(n - \hat{s}_2)}.$$

Since  $n - \hat{s}_2 \geq n_2$  (recall that  $n_1 + n_2 = n$  and  $n_1 \geq \hat{s}_2$  from inequality (11)), the above inequality implies that  $q_1^a(n - \hat{s}_1) > 1/2$ . By the definition of  $\hat{s}_1$  we know that  $q_1^b(\hat{s}_1) \geq 1/2$ , but then  $q_1^a(n - \hat{s}_1) > 1/2$  contradicts equation (12).

Now, we complete the proof by showing that (13) holds. Let  $P^a(n', k)$  denote the probability that, in a society with  $n^1$  voters with  $p_1$  and  $n'$  voters with  $p_2$ , exactly  $k$  of the voters support  $a$ . So in this calculation, the number of voters of type 1 is always fixed, but the number of voters of type 2 is given by  $n'$ . Writing in the expressions for  $q_2^a$  from Bayes' rule, we need to show that

$$\frac{1}{k} \left( \frac{(1 - p_2)P^a(n^2 - 1, k - 1)}{p_2 P^a(n^2 - 1, k) + (1 - p_2)P^a(n^2 - 1, k - 1)} \right) >$$

$$\frac{1}{k+1} \left( \frac{(1-p_2)P^a(n^2-1, k)}{p_2P^a(n^2-1, k+1) + (1-p_2)P^a(n^2-1, k)} \right). \quad (14)$$

Note that

$$P^a(n', k) = p_2P^a(n'-1, k) + (1-p_2)P^a(n'-1, k-1). \quad (15)$$

So substituting from (15) and simplifying, we rewrite (14) as

$$(k+1) \frac{P^a(n^2-1, k-1)}{P^a(n^2, k)} > k \frac{P^a(n^2-1, k)}{P^a(n^2, k+1)}. \quad (16)$$

We show this by induction on  $n^2$ . A straightforward (but tedious) expansion of the expressions (that we leave to the reader) verifies that (16) holds for  $n_2 = 1$  and any  $k \geq 1$  (set  $\frac{P^a(n^2, k)}{P^a(n^2, k+1)} = 0/0 = 1$  when  $k > n^1 + n^2$ ). We now show that if (16) holds for each  $n^2 < n'$  and  $k \geq 1$ , then it holds for  $n'$  and any  $k \geq 1$ . Rewriting (16) at  $n'$  and expanding using (15) in each expression we obtain,

$$\begin{aligned} & (k+1) \left( p_2^2 P^a(n'-2, k-1) P^a(n'-1, k+1) + p_2(1-p_2) P^a(n'-2, k-1) P^a(n'-1, k) \right. \\ & \left. + p_2(1-p_2) P^a(n'-2, k-2) P^a(n'-1, k+1) + (1-p_2)^2 P^a(n'-2, k-2) P^a(n'-1, k) \right) \\ & > k \left( p_2^2 P^a(n'-2, k) P^a(n'-1, k) + p_2(1-p_2) P^a(n'-2, k) P^a(n'-1, k-1) \right. \\ & \left. + p_2(1-p_2) P^a(n'-2, k-1) P^a(n'-1, k) + (1-p_2)^2 P^a(n'-2, k-1) P^a(n'-1, k-1) \right) \end{aligned}$$

Using the induction hypothesis, we eliminate the first expression on each side of the inequality, and then collecting terms and simplifying we obtain

$$\begin{aligned} & p_2 P^a(n'-2, k-1) P^a(n'-1, k) + (k+1) p_2 P^a(n'-2, k-2) P^a(n'-1, k+1) \\ & \quad + (1-p_2) P^a(n'-2, k-2) P^a(n'-1, k) \\ & > k p_2 P^a(n'-2, k) P^a(n'-1, k-1) + (1-p_2) P^a(n'-2, k-1) P^a(n'-1, k-1) \end{aligned}$$

Now, substituting for  $P^a(n'-1, \cdot)$  from (15), we rewrite the above as

$$\begin{aligned} & p_2^2 P^a(n'-2, k-1) P^a(n'-2, k) + p_2(1-p_2) P^a(n'-2, k-1) P^a(n'-2, k-1) \\ & \quad + p_2(1-p_2) P^a(n'-2, k-2) P^a(n'-2, k) + (1-p_2)^2 P^a(n'-2, k-2) P^a(n'-2, k-1) \\ & + (k+1) p_2^2 P^a(n'-2, k-2) P^a(n'-2, k+1) + (k+1) p_2(1-p_2) P^a(n'-2, k-2) P^a(n'-2, k) \\ & > p_2(1-p_2) P^a(n'-2, k-1) P^a(n'-2, k-1) + (1-p_2)^2 P^a(n'-2, k-2) P^a(n'-2, k-1) \\ & \quad + k p_2^2 P^a(n'-2, k) P^a(n'-2, k-1) + k p_2(1-p_2) P^a(n'-2, k) P^a(n'-2, k-2) \end{aligned}$$

Simplifying, we must only show the inequality

$$(k+1) p_2 P^a(n'-2, k-2) P^a(n'-2, k+1) + (k+1) (1-p_2) P^a(n'-2, k-2) P^a(n'-2, k)$$

$$> (k-1)p_2P^a(n'-2, k)P^a(n'-2, k-1) + (k-1)(1-p_2)P^a(n'-2, k)P^a(n'-2, k-2)$$

Using (15) at  $n' - 1$  we rewrite this as

$$(k+1)P^a(n'-2, k-2)P^a(n'-1, k+1) > (k-1)P^a(n'-2, k)P^a(n'-1, k-1). \quad (17)$$

So we need only show that (17) holds. By the induction hypothesis, we know that

$$(k+1)P^a(n'-2, k-1)P^a(n'-1, k+1) > kP^a(n'-2, k)P^a(n'-1, k),$$

and

$$kP^a(n'-2, k-2)P^a(n'-1, k) > (k-1)P^a(n'-2, k-1)P^a(n'-1, k-1),$$

or

$$kP^a(n'-1, k) > (k-1)P^a(n'-2, k-1)P^a(n'-1, k-1)/P^a(n'-2, k-2).$$

Combined, these imply that

$$\begin{aligned} & (k+1)P^a(n'-2, k-1)P^a(n'-1, k+1) \\ & > P^a(n'-2, k)(k-1)P^a(n'-2, k-1)P^a(n'-1, k-1)/P^a(n'-2, k-2), \end{aligned}$$

which simplifies to

$$(k+1)P^a(n'-1, k+1)P^a(n'-2, k-2) > (k-1)P^a(n'-2, k)P^a(n'-1, k-1).$$

This verifies that (17) holds and completes the proof. ■

**Proof of Theorem 5:** First, note that given the single-peaked preferences (accounting for the possibility of two peaks),  $(s^{\text{maj}}, S)$  is self-stable if and only if

$$|\{i : U_i(s^{\text{maj}}) \geq U_i(s^{\text{maj}} - 1)\}| > n - S \quad \text{and} \quad |\{i : U_i(s^{\text{maj}}) \geq U_i(s^{\text{maj}} + 1)\}| > n - S. \quad (18)$$

A sufficient condition for this is that

$$|\{i : U_i(s^{\text{maj}}) \geq U_i(s^{\text{maj}} - 1)\}| > n - S \quad \text{and} \quad |\{i : U_i(s^{\text{maj}} - 1) \geq U_i(s^{\text{maj}})\}| > n - S.$$

which is in turn guaranteed by

$$S > |\{i : U_i(s^{\text{maj}} - 1) \geq U_i(s^{\text{maj}})\}| > n - S. \quad (19)$$

Recall from (2) that

$$U_i(s) - U_i(s-1) = (1-p_i)P_i(s-1) - p_iP_i(s-2).$$

Thus,

$$\{i : U_i(s^{\text{maj}} - 1) \geq U_i(s^{\text{maj}})\} = \left\{ i : \frac{p_i}{1-p_i} = z_i \geq \frac{P_i(s^{\text{maj}} - 1)}{P_i(s^{\text{maj}} - 2)} \right\}. \quad (20)$$

From the definition of  $P_i(s)$  it follows that

$$\frac{P_i(s)}{P_i(s-1)} = \frac{\sum_{C \subset N \setminus i, |C|=s} [\times_{j \in C} p_j \times_{j \notin C} (1-p_j)]}{\sum_{C \subset N \setminus i, |C|=s-1} [\times_{j \in C} p_j \times_{j \notin C} (1-p_j)]}$$

Dividing top and bottom by  $\times_{j \neq i} (1-p_j)$ , this becomes

$$\frac{P_i(s)}{P_i(s-1)} = \frac{\sum_{C \subset N \setminus i, |C|=s} \times_{j \in C} z_j}{\sum_{C \subset N \setminus i, |C|=s-1} \times_{j \in C} z_j}$$

So, by the above equation and (20), we can rewrite (19) as

$$S > |\{i : z_i \geq \frac{\sum_{C \subset N \setminus i, |C|=\frac{n}{2}} \times_{j \in C} z_j}{\sum_{C \subset N \setminus i, |C|=\frac{n}{2}-1} \times_{j \in C} z_j}\}| > n - S. \quad (21)$$

This can be rewritten as

$$S > |\{i : \sum_{C \subset N, i \in C, |C|=\frac{n}{2}} \times_{j \in C} z_j \geq \sum_{C \subset N \setminus i, |C|=\frac{n}{2}} \times_{j \in C} z_j\}| > n - S,$$

which is the claimed expression.

A direct rewriting of (21) leads to the claimed expression in (4):

$$S > |\{i : z_i \geq \sum_{k \neq i} \lambda_k^i z_k\}| > n - S,$$

where

$$\lambda_k^i = \frac{2 \sum_{|C|=\frac{n}{2}-1, i, k \notin C} (\times_{j \in C} z_j)}{n \sum_{|C|=\frac{n}{2}-1, i \notin C} (\times_{j \in C} z_j)}.$$

Direct inspection shows that  $\sum_{k \neq i} \lambda_k^i = 1$  for all  $i$ . ■