OPTIMAL PRIMARIES

RICHARD HOLDEN* AND PATRICK HUMMEL**

ABSTRACT. We analyze a model of US presidential primary elections for a given political party. There are two candidates, one of whom is a higher quality candidate. Voters reside in $m$ different states and receive noisy private information about the identity of the superior candidate. States vote in some order, and this order is chosen by a social planner. Voters draw rational inferences about the identity of the superior candidate from both their own signals and the results of primaries in earlier states. We provide conditions under which the ordering of the states that maximizes the probability that the higher quality candidate is elected is for states to vote in order from smallest to largest populations and most accurate private information to least accurate private information.

1. INTRODUCTION

A striking and unusual feature of American politics is the presidential primary election system. This system involves elections by 50 states, where the states vote in some predetermined order over the course of several months. The order of the primaries is, for a given party, largely determined by the national party in question. As this nomination system has played a crucial role in determining the president of the United States, it seems important to understand how to design an optimal presidential primary. In this paper, we make a contribution towards this question.

Most people involved in professional politics believe that the order of the states in the primaries matters. Participants seem unanimous in the view that strong results in early states create momentum and lead to an information cascade whereby voters in later states ...
become much more likely to vote for candidates who were successful in early states. Theoretical models such as Ali and Kartik (2010) show that such behavior has a rational basis.\(^2\) Moreover, there is significant empirical support that these momentum effects play an important role in elections (e.g. Bartels (1985, 1988), Kenney and Rice (1994), Knight and Schiff (2010), and Popkin (1994)). For example, Bill Clinton swept to the nomination from seeming obscurity in 1992 as the result of an unexpectedly strong finish in New Hampshire, and John Kerry demolished the field in 2004 after wins in New Hampshire and Iowa.

The possibility of momentum in presidential primaries leads many states to seek opportunities to have an early influence on the campaign. Residents of Iowa and New Hampshire cherish their early role—sometimes being described as “the presidential wine-tasters of America”.\(^3\) Mayer and Busch (2004) note that since 1988 states have been engaging in a process known as “front-loading”, in which states attempt to hold their elections earlier and earlier in the campaign season. As a result, roughly 70 percent of delegates are decided by March 2 today compared to 10 percent in 1976 (Redslawk et al., 2011). And in the 2008 presidential primary, a number of states sought to hold their primaries so early that these states were forced to have half of their delegates to the nominating convention stripped as a consequence.

Candidates also relish the opportunity to try to capture victories in early states. Brams and Davis (1982) note that candidates in the 1976 and 1980 US presidential primaries heavily emphasized spending in states that held their primaries early. Norrander (1992) notes that even the introduction of Super Tuesday and the large number of votes it made up for grabs on a single day was not sufficient to deter candidates from focusing their resources on the early states. And the emphasis candidates place on early states has persisted in the most recent presidential primaries. For instance, in the 2008 presidential primary, Barack Obama focused the bulk of his resources to New Hampshire, John Edwards did the same in Iowa, and Chris Dodd moved his family to Iowa two years prior.

\(^2\)In addition, classic papers on social learning such as Bikhchandani, Hirshleifer and Welch (1992) and Banerjee (1992) have shown that information cascades can be rational in a variety of other economic contexts.

\(^3\)As quoted in *The West Wing.*
It is not only the participants who should care about the order of the primaries. Each major party wants to field the strongest possible candidate in the general election, and the order of the states can affect whether that candidate is the nominee. And voters in aggregate want to elect the best possible president, and hence should also care.

Moreover, there has been substantial criticism of the existing order of primaries. As the New York Times editorial page put it in September 2008: “The presidential primary system is broken. For years, the nominating process has unfolded in an orderly, if essentially unfair, way. The schedule has worked very nicely for early-voting states, which have had a steady stream of would-be presidents knocking on their doors, making commitments on issues like the Iowa full-employment program, also known as the ethanol subsidy. The losers have been states like New York and California, which have often gotten to vote only when the contests were all but decided. Issues that matter to them, like mass transportation, have suffered.”

In this paper we take a mechanism design approach to the question of the optimal order of primaries. One may view our mechanism designer either as a self-interested political party who wants to nominate the best possible candidate, or a benevolent social planner who wants to maximize the total utility of the voters in the primary. Voters in our model draw rational inferences about the identity of the superior candidate from the results of primaries in early states, and this in turn can lead to momentum and information cascades. The planner seeks to design a primary that maximizes the probability that the higher quality candidate is elected when the order of the states may affect the information cascades that take place.

Our model allows for heterogeneity at the individual, state, and national levels. Individual voters may differ in their own private preferences by having biased preferences for a candidate. States may differ in their size, the precision of the information their voters receive about the candidates, and how representative their voters are of the country as a whole. And nationally voters as a whole will have different preferences depending on which candidate is the higher quality candidate.
We derive conditions under which a social planner can increase the probability that the higher quality candidate is elected by ordering the states from smallest to largest (i.e. allowing the smallest state to vote first and the largest state to vote last). We also show that allowing states that receive more precise signals to vote earlier typically increases the probability that the higher quality candidate is elected. Finally, we discuss the robustness of the results when candidate strategies are an endogenous function of the order in which the states vote.

The results are significant because they suggest a possible rationalization for why one might wish to allow small states such as Iowa and New Hampshire to vote early in a presidential primary. Ordering the states from smallest to largest can increase the probability that the best candidate is elected. Similarly, the result that it is advantageous to allow states with more precise signals to vote earlier also suggests a reason why it may be advantageous to allow states such as Iowa and New Hampshire to vote early. For one, better educated voters tend to have more well-informed opinions about the abilities of the candidates, and US Census Data from the past few decades reveals that the percentage of residents over the age of 25 in Iowa and New Hampshire who have graduated from high school is consistently among the few highest of all the states in the country.\(^4\)

Furthermore, when candidates campaign in small states, voters are more likely to be able to meet the candidates individually and obtain precise signals about their quality. Roughly two months before the 2008 Iowa caucus, nearly two-thirds of voters in Iowa had already had an opportunity to meet a candidate.\(^5\) And a few weeks before the 1996 New Hampshire Republican primary, approximately 20 percent of people polled by a WMUR-Dartmouth College survey said they had met or seen in person a candidate in the primary (Buhr 2000). By contrast, in a large state such as California, it would be infeasible for such a large percentage of voters to meet a candidate before the election.\(^6\) Vavrek et al. (2002)


\(^5\)This was noted in a November 2007 CBS News/New York Times Poll.

\(^6\)Redslawk et al. (2011) attribute the fact that so many voters in Iowa are able to meet the candidates in person to Iowa’s small size and Vavrek et al. (2002) claim that New Hampshire’s small size creates ideal opportunities for voters to interact with the candidates.
provide both reduced form and causal evidence that such meetings enable voters to receive precise information about the candidates, stating: “Meeting the candidates face-to-face, receiving direct mail, and getting phone calls on behalf of candidates all have systematic effects on voters’ uncertainty, knowledge, and attitudes about candidates. Voters’ personal interactions with candidates are most important in reducing their uncertainty about how to rate candidates.”

2. The Model

2.1. Statement of the Problem. There are two candidates, A and B, and m states $S_1, \ldots, S_m$. Each state $S_j$ has a continuum of voters of measure $\lambda_j$, and the election is decided by majority rule. Thus if $y_j$ denotes the fraction of voters that votes for candidate A in state $S_j$, then candidate A is elected if $\sum_{j=1}^m y_j \lambda_j > \frac{1}{2} \sum_{j=1}^m \lambda_j$, both candidates are elected with probability $\frac{1}{2}$ if $\sum_{j=1}^m y_j \lambda_j = \frac{1}{2} \sum_{j=1}^m \lambda_j$, and candidate B is elected otherwise.

There are two states of the world, $a$ and $b$. If the state of the world is $a$, then candidate A is the higher quality candidate, and if the state of the world is $b$, then candidate B is the higher quality candidate. No voter observes the state of the world directly, but each voter shares a common prior that the probability the state of the world is $a$ is $\pi$ for some $\pi \in (0, 1)$.

If the state of the world is $a$, then each voter in state $S_j$ receives a private signal that is an independent and identically distributed draw from a distribution that takes on the value $\alpha$ with probability $p_j > \frac{1}{2}$ and the value $\beta$ with probability $1 - p_j$. And if the state of the world is $b$, then each voter in state $S_j$ receives a private signal that is an independent and identically distributed draw from a distribution that takes on the value $\beta$ with probability $p_j > \frac{1}{2}$ and the value $\alpha$ with probability $1 - p_j$. We assume that $p_j > \max\{\pi, 1 - \pi\}$ for all $j$ to ensure that a voter will initially believe it is more likely that candidate A is a better

---

$^7$Jones (1998) and Shaw (1999) also offer evidence that the number of visits to a state is related to total votes received.
candidate after receiving the signal $\alpha$ and candidate $B$ is a better candidate after receiving the signal $\beta$.

There are three possible types of voters in the population. The first possibility is that a voter is an $A$-partisan and always obtains utility 1 from electing candidate $A$ and utility 0 from electing candidate $B$. A second possibility is that a voter is a $B$-partisan and always obtains utility 1 from electing candidate $B$ and utility 0 from electing candidate $A$. And a last possibility is that a voter is a neutral. A neutral obtains utility 1 from electing candidate $A$ in state $a$, utility 0 from electing candidate $B$ in state $a$, utility 1 from electing candidate $B$ in state $b$, and utility 0 from electing candidate $A$ in state $b$. Throughout we let $\pi^j_A$ denote the fraction of voters in state $S_j$ who are $A$-partisans, $\pi^j_B$ denote the fraction of voters in state $S_j$ who are $B$-partisans, and $\pi^j_N = 1 - \pi^j_A - \pi^j_B$ denote the fraction of voters in state $S_j$ who are neutrals.

While voters in a given state $S_j$ all vote at the same time, different states vote at different times, and some states vote after observing how voters in other states have voted. We assume that all voters in state $S_{(1)}$ vote without observing how voters in other states have voted, and that voters in state $S_{(j)}$ vote after observing how voters in states $S_{(1)}, \ldots, S_{(j-1)}$ have voted. Throughout we also assume that all voters vote sincerely. Thus $A$-partisans vote for candidate $A$, $B$-partisans vote for candidate $B$, and a neutral voter votes for candidate $A$ if and only if this voter believes the probability the state of the world is $a$ is greater than $\frac{1}{2}$ at the time the voter votes.

We consider the problem of a social planner who wishes to select the order in which the states vote, $S_{(1)}, \ldots, S_{(m)}$, to maximize the probability that the higher quality candidate is elected. While the planner and the voters know the values of $\pi^j_N$, we assume that the planner and the voters do not know the values of $\pi^j_A$ and $\pi^j_B$ at the time that the planner chooses the order of the states. Instead the planner and the voters only know that each value of $\pi^j_A$ is an independent draw from a continuous cumulative distribution function $F_j$ with support equal to some subset of the interval $[0, 1 - \pi^j_N]$ and corresponding density $f_j$.

\textsuperscript{8}Selman (2010) analyzes a model of optimal primaries that differs from ours in that the values of $\pi^j_A$ and $\pi^j_B$ in each state are common knowledge. In such a model, voters can deduce the identity of the superior candidate
2.2. Discussion of the Model. Before proceeding to the analysis, we pause briefly to discuss some of our modeling assumptions. The most notable simplifications are the restriction to elections with two candidates and the fact that we allow for only three types of voters. The restriction to elections with two candidates is purely for simplicity, and we do not believe that it is consequential for the results. In particular, presidential primaries with multiple candidates typically narrow down to the two most serious candidates very quickly (Abramson et al., 1992; Aldrich, 1980; Bartels, 1988; Matthews, 1978; Popkin, 1994), so the analysis we present is quite relevant for the critical final phase of presidential primaries with multiple candidates. And the assumption that a voter obtains utility 1 from the election of her preferred candidate and 0 from her less preferred candidate has no effect on the analysis since, for example, if the $A$-partisans differed in the intensities of their preferences over the two candidates, such voters would still vote the same way as before.

Finally, it is worth noting our reasons for working with these assumptions about voter types rather than another logical approach. While we have assumed that each voter’s utility either depends only on a private value component or only on a common value component, another possible modeling approach would be to assume that each voter’s utility difference between the candidates depends both on a private value component and on a common value component. However, we feel that this alternative modeling approach is less appropriate for addressing questions related to the optimal order of the states. Under this alternative approach, the cases in which the order of the states is most likely to affect the outcome of the election are the cases in which the election is likely to be close or the cases in which most voters prefer one candidate on the basis of private values but the other candidate on the basis of common values. In such cases, it is not clear which candidate would be better for a benevolent social planner to elect. By contrast, in the present model, the majority of voters prefer the higher quality candidate if most voters are neutrals and it makes sense for a social planner to strive to maximize the probability of electing the higher quality candidate.

with virtual certainty after the election in the first state. By contrast, in our model, the uncertainty about the values of $\pi_A^1$ and $\pi_B^1$ means that there may be uncertainty about the identity of the superior candidate even after several states hold their elections.
3. Results

Suppose that all neutral voters in state $S_j$ who receive a private signal $\alpha$ vote for $A$ and all neutral voters in state $S_j$ who receive a private signal $\beta$ vote for $B$. Then if the state of the world is $a$, the fraction of voters that votes for candidate $A$ in state $S_j$ is $\pi_A^j + p_j \pi_N^j$, and if the state of the world is $b$, the fraction of voters that votes for candidate $A$ in state $S_j$ is $\pi_A^j + (1 - p_j) \pi_N^j$.

Thus if $y_j$ denotes the fraction of voters that votes for candidate $A$ in a state $S_j$ where all neutral voters vote according to their private signals, then either the state of the world is $a$ and $\pi_A^j = y_j - p_j \pi_N^j$ or the state of the world is $b$ and $\pi_A^j = y_j - (1 - p_j) \pi_N^j$. From this it follows that if all neutral voters in states $S_{(1)}, \ldots, S_{(j)}$ vote according to their private signals and $\pi(j)$ denotes the probability with which an outside observer believes the state of the world is $a$ after states $S_{(1)}, \ldots, S_{(j)}$ have voted, then

$$\pi(j) = \frac{\pi(j-1) f(j-1)(y_{(j-1)} - p_{(j-1)} \pi_N^{(j-1)})}{\pi(j-1) f(j-1)(y_{(j-1)} - p_{(j-1)} \pi_N^{(j-1)}) + (1 - \pi(j-1)) f(j-1)(y_{(j-1)} - (1 - p_{j-1}) \pi_N^{(j-1)})}.$$

Now we specialize to the case in which $\pi_N^j = \pi_N$ for all $j$ and $\pi_A^j$ is drawn from the uniform distribution on $[0, 1 - \pi_N]$ for all $j$. Thus $f_j(x) = \frac{1}{1 - \pi_N}$ for all $x \in [0, 1 - \pi_N]$ and $f_j(x) = 0$ for all $x \not\in [0, 1 - \pi_N]$. Under these assumptions, if all neutral voters in state $S_{(j)}$ vote according to their private signals and $y_{(j)} > 1 - p_{(j)} \pi_N$, then $\pi_{(j)} = 1$, as the only way so many voters could have voted for candidate $A$ in state $S_{(j)}$ is if the state of the world is $a$. If $y_{(j)} < p_{(j)} \pi_N$, then $\pi_{(j)} = 0$, as the only way so few voters could have voted for candidate $A$ in state $S_{(j)}$ is if the state of the world is $b$. But if $p_{(j)} \pi_N \leq y_{(j)} \leq 1 - p_{(j)} \pi_N$, then $f_{(j)}(y_{(j)} - p_{(j)} \pi_N^{(j)}) = f_{(j)}(y_{(j)} - (1 - p_{(j)}) \pi_N^{(j)})$ and $\pi_{(j)} = \pi_{(j-1)}$.

Thus under these assumptions, neutral voters in state $S_{(j)}$ follow simple strategies in deciding which candidate to vote for. If a candidate has received a fraction of the vote greater than $1 - p_{(k)} \pi_N$ in some previous state $S_{(k)}$ with $k < j$, then voters recognize that this candidate is the higher quality candidate, and all neutral voters in state $S_{(j)}$ vote for that particular candidate. If no candidate has received a fraction of the vote greater than
1 - p(k)\pi_N in some previous state \( S(k) \) with \( k < j \), then all neutral voters in state \( S(j) \) vote according to their private signals. We use this insight in proving the following proposition:

**Proposition 1.** Suppose that \( \pi^j_N = \pi_N \) for all \( j \), \( \pi^j_A \) is drawn from the uniform distribution on \([0, 1 - \pi_N]\) for all \( j \), and \( p_j = p > \frac{1}{2} \) for all \( j \). Then the ordering of the states that maximizes the probability that the higher quality candidate wins the election is the ordering in which the states are ordered from smallest to largest, or the ordering in which state \( S(1) \) has the smallest value of \( \lambda_j \), state \( S(2) \) has the second smallest value of \( \lambda_j \), and in general state \( S(k) \) has the \( k^{th} \) smallest value of \( \lambda_j \).

The intuition for this result is as follows. Note that if voters are able to deduce the identity of the higher quality candidate from the results of an early state, then more voters will have an opportunity to act on this information if a larger number of voters remain to vote, and more voters will vote for the higher quality candidate if larger states remain to vote. At the same time, if voters are not able to deduce the identity of the higher quality candidate from the results of an early state, then the order of the states will never affect the number of voters who vote for the higher quality candidate.

Thus under either scenario, at least as many voters vote for the higher quality candidate if the small states vote before the big states, and the higher quality candidate wins with the greatest possible probability if the states are ordered from smallest to largest. The result follows formally by showing that the distribution of the number of voters who vote for the higher quality candidate when the states are ordered from smallest to largest first order stochastically dominates this distribution when the states are ordered in some other way.

One can further illustrate how the optimal order of the states depends on the accuracy of each voter’s private information in each state. This is done in Proposition 2:

**Proposition 2.** Suppose that \( \pi^j_N = \pi_N \) for all \( j \), \( \pi^j_A \) is drawn from the uniform distribution on \([0, 1 - \pi_N]\) for all \( j \), and \( \lambda_j = \lambda \) for all \( j \). Then the ordering of the states that maximizes the probability that the higher quality candidate wins the election is the ordering in which the states are ordered from the most informative private signals to the least informative private
signals, or the ordering in which state $S_{(1)}$ has the largest value of $p_j$, state $S_{(2)}$ has the second largest value of $p_j$, and in general state $S_{(k)}$ has the $k^{th}$ largest value of $p_j$.

To understand why this result holds, note that the result of an election in a state where voters have accurate private signals is relatively more likely to give information that reveals the identity of the higher quality candidate than the result of an election in a state where voters have less accurate private signals. Thus if states where voters have more accurate private information vote first, then the identity of the higher quality candidate will normally be revealed more quickly than if states where voters have less accurate private information vote first. For this reason, a larger number of states will have opportunities to take the known identity of the higher quality candidate into account when voting if states with more accurate private signals vote first, and there will be more later states where all neutral voters vote for the higher quality candidate if states with more accurate private signals vote first.

At the same time, more voters in early states will vote for the higher quality candidate when states where voters have accurate private information vote first because voters in early states vote according to their private information. Thus allowing the states with the most accurate private information to vote first increases both the number of voters who vote for the higher quality candidate in early states and the number of voters who vote for the higher quality candidate in later states. From this it follows that the higher quality candidate wins with greater probability if the states with the most accurate private signals vote first. The formal proof of the result again illustrates that the distribution of the number of voters who vote for the higher quality candidate when the states are ordered from most accurate signals to least accurate signals first order stochastically dominates this distribution when the states are ordered in some other way.

Propositions 1 and 2 have both made use of the assumption that the fraction of partisans is drawn from a uniform distribution. However, the assumption that the fraction of partisans is drawn from a uniform distribution rules out the possibility that a bandwagon may begin in a state that benefits the lower quality candidate. Since these bad bandwagons may take place under more general distributions, it is natural to ask how the results can be extended
without this assumption. We thus seek to illustrate that the results in Propositions 1 and 2 can be extended even if a bandwagon may start in a later state that helps the lower quality candidate. As this is difficult to analyze under the general case with a large number of states, we assume there are \( m = 2 \) states in analyzing this possibility.

To address this, we consider what happens when the fraction of \( A \)-partisans in a given state is drawn from a distribution with a density \( f \) such that this fraction is likely to assume moderate values (those between \( \delta \) and \( 1 - \pi_N - \delta \) for some \( \delta \in (0, \frac{1-\pi_N}{2}) \)) and relatively unlikely to assume extreme values (those less than \( \delta \) or greater than \( 1 - \pi_N - \delta \)). Such a density will allow for the possibility of bad bandwagons, as if the fraction of \( A \)-partisans in a given state happens to be less than \( \delta \), then voters in later states will conclude that it is relatively more likely that \( B \) was the better candidate and vote for \( B \) even if \( A \) is actually the better candidate. Moreover, this density will continue to have the natural feature that neutral voters in the second state will vote for \( A \) if \( A \) does sufficiently well in the first state, vote for \( B \) if \( B \) does sufficiently well in the first state, and vote according to their private signal for more moderate outcomes of the election in the first state.

Formally, we will assume

**Condition 1.** \( f(x) \) is symmetric and weakly single-peaked about \( x = \frac{1-\pi_N}{2} \) and there exists some positive \( \delta \) satisfying \( \delta < \max\{\pi_N(2p_j - 1), \frac{1-2p_j\pi_N}{2}\} \) for \( j = 1 \) and 2 such that

- \( f(x) \) is constant for \( x \in [\delta, 1 - \pi_N - \delta] \),
- \( \frac{\pi f(x)}{\pi f(x)+(1-\pi)f(z)} < \min\{1-p_1, 1-p_2\} \) if \( x < \delta \) and \( z \in [\delta, 1 - \pi_N - \delta] \),
- \( \frac{\pi f(x)}{\pi f(x)+(1-\pi)f(z)} > \max\{p_1, p_2\} \) if \( x \in [\delta, 1 - \pi_N - \delta] \) and \( z > 1 - \pi_N - \delta \) and
- \( \max\{1-p_1, 1-p_2\} < \frac{\pi f(x)}{\pi f(x)+(1-\pi)f(z)} < \min\{p_1, p_2\} \) for all other values of \( x \in [0, 1-\pi_N] \) and \( z \in [0, 1 - \pi_N] \).

Under such a density \( f(x) \), the fraction of \( A \)-partisans is considerably more likely to take on a particular value between \( \delta \) and \( 1 - \pi_N - \delta \) than a particular value less than \( \delta \) or greater than \( 1 - \pi_N - \delta \). Moreover, if a candidate does poorly in the first state because the fraction of partisans for the candidate in that state is less than \( \delta \), then voters conclude that candidate
is unlikely to be the better candidate and neutral voters in the second state all vote for the other candidate. But if both candidates do moderately well in the first state, then voters still think there is a substantial chance that either candidate could be the better candidate and voters continue to act on their private information in the second state. Under these assumptions we obtain the following result:

**Proposition 3.** Suppose that \( m = 2 \), \( \pi_N^1 = \pi_N \), \( p_j = p \), and \( f_j = f \) for all \( j \), \( \pi_N \geq \frac{1}{2} \), and Condition 1 holds. Then the ordering of the states that maximizes the probability that the higher quality candidate wins the election is the ordering in which the smaller state votes first and the larger state votes last.

To understand the intuition for this result, note that if there are two states and there is a bandwagon for candidate A, then candidate A wins the election regardless of the order of the states. In order for the first state to induce a bandwagon for candidate A, it is necessary that at least half of the voters vote for A in the first state. And if there is a bandwagon for candidate A, then at least half of the voters vote for A in the second state. Thus if there is a bandwagon for candidate A, then the majority of voters vote for A and A wins the election regardless of the order of the states. Similarly, if there is a bandwagon for candidate B, then candidate B wins the election regardless of the order of the states.

Thus the only way the order of the states can affect which candidate is elected is if there is no bandwagon for either candidate. Now if there is no bandwagon for either candidate, then both candidates obtain a moderate number of votes in the first state, and the fraction of votes received by the higher quality candidate in the first state is necessarily close to \( \frac{1}{2} \). At the same time, on average the majority of voters will vote for the higher quality candidate in the second state because the majority of neutral voters believe the higher quality candidate is the better candidate.

Thus if there is no bandwagon and the large state votes first, then the fraction of voters that votes for the higher quality candidate in the large state is close to \( \frac{1}{2} \) and the majority of voters are expected to vote for the higher quality candidate in the small state. And if there is no bandwagon and the small state votes first, then the fraction of voters that votes
for the higher quality candidate in the small state is close to \( \frac{1}{2} \) and the majority of voters are expected to vote for the higher quality candidate in the large state. This indicates that more voters will be expected to vote for the higher quality candidate in the scenario in which the small state votes first than in the scenario in which the large state votes first. For this reason the higher quality candidate wins with greater probability when the small state votes first.

One can also extend the substantive conclusions of Proposition 2 to this framework that allows for the possibility of bad bandwagons. This is done in Proposition 4:

**Proposition 4.** Suppose that \( m = 2, \pi^j_N = \pi_N, \lambda_j = \lambda, \) and \( f_j = f \) for all \( j \), and Condition 1 holds. Then the ordering of the states that maximizes the probability that the higher quality candidate wins the election is the ordering in which the state with more informative private signals votes first and the state with less informative private signals votes last.

To understand why this result holds, note that if the state with the more accurate private signals votes first, then it is relatively more likely that there will be a bandwagon for the higher quality candidate because the state is relatively more likely to reveal accurate information that the higher quality candidate is the better candidate. It is also relatively less likely that there will be a bandwagon for the lower quality candidate because the state is relatively less likely to reveal inaccurate information that the lower quality candidate is the better candidate. Thus relatively more favorable types of bandwagons take place when the state with the more accurate private signals votes first.

At the same time, conditional on a bandwagon taking place, one would prefer that the state with the more accurate private signals votes first. If a bandwagon takes place, then voters in the second state do not condition their votes on their signals, and such voters vote the same way regardless of the accuracy of their signals. But more voters in the first state vote for the higher quality candidate if the voters in the first state have relatively more accurate private signals. Thus if a bandwagon takes place, more voters vote for the higher quality candidate if the state with the more accurate private signals votes first. And if no
bandwagon takes place, then the number of voters who vote for the higher quality candidate is not affected by the order of the states.

Combining the ideas in these two paragraphs indicates that having the state with the more accurate private signals vote first can only increase the number of voters that votes for the higher quality candidate. Thus the higher quality candidate wins with greater probability if the state with the more accurate private signals votes first.

4. ENDOWENOUS CANDIDATE STRATEGIES

In the previous section we have assumed that changing the order of the states does not affect the distribution of voter preferences in any particular state. This assumption might seem questionable since candidates may wish to modify where they focus their campaign resources if the order of these states is changed, and this may in turn change the distribution of voter preferences. Thus we address the question of how allowing for endogenous candidate strategies affects the main results of the paper in this section.

Throughout this section we assume that by spending more money in a given state, a candidate can increase the fraction of partisan supporters for that candidate in the state while simultaneously decreasing the fraction of partisan supporters for the other candidate by the same amount. In particular, if candidate $A$ wishes to increase the fraction of $A$-partisans in state $S_j$ by $a_j$, then candidate $A$ must spend some amount $c_j(a_j)$ in state $S_j$, where $c_j(a)$ is a strictly increasing function of $a$ satisfying $c_j(0) = 0$. Similarly, if candidate $B$ wishes to increase the fraction of $B$-partisans in state $S_j$ by $b_j$, then candidate $B$ must spend $c_j(b_j)$ in state $S_j$ for the same function $c_j$.

Formally, the strategies for the candidates are as follows. Candidate $A$ chooses an allocation $a = (a_1, a_2, \ldots, a_m) \geq 0$ satisfying $\sum_{j=1}^{m} c_j(a_j) = C$ while candidate $B$ simultaneously chooses an allocation $b = (b_1, b_2, \ldots, b_m) \geq 0$ satisfying $\sum_{j=1}^{m} c_j(b_j) = C$, where $C > 0$ is the budget available to a candidate. Each candidate’s objective is to maximize the probability that he or she is elected. Throughout we assume that each candidate believes that he or she is the higher quality candidate with probability $\pi = \frac{1}{2}$ and that $\pi_A^j$ is drawn from a distribution
with density $f_j$ that is symmetric about $(1 - \pi^j_N)/2$ in the sense that $f_j(x) = f_j(1 - \pi^j_N - x)$ for all $x \in [0, 1 - \pi^j_N]$.

If candidate $A$ chooses the allocation $a = (a_1, a_2, \ldots, a_m)$, candidate $B$ chooses the allocation $b = (b_1, b_2, \ldots, b_m)$, and $\pi^j_A$ is the original draw of the fraction of $A$-partisans in state $S_j$ from the distribution $f_j$, then we assume that the fraction of $A$-partisans in state $S_j$ is $\pi^j_j(\pi^j_A, a_j, b_j)$, where $\pi^j_j(\pi^j_A, a_j, b_j)$ is $\pi^j_A + a_j - b_j$ if $\pi^j_A + a_j - b_j \in [0, 1 - \pi^j_N]$, 0 if $\pi^j_A + a_j - b_j < 0$, and $1 - \pi^j_N$ if $\pi^j_A + a_j - b_j > 1 - \pi^j_N$. The fraction of neutral voters in state $S_j$ remains unchanged regardless of the candidates’ choices of strategies. Voters then vote sincerely after candidates choose their campaign strategies.

Given these assumptions we obtain the following result:

**Proposition 5.** Suppose there exists a pure strategy equilibrium to the candidate budget allocation game. Then the distribution of voter preferences in each state is the same as in the original model without endogenous candidate strategies.

The economic intuition for this result is as follows. While candidates may wish to change how they allocate their resources to the various states if the order of the states is changed, for any fixed order of the states, candidates are likely to have similar beliefs about which states are the most important states to campaign in. For instance, candidates may always think that the earliest states are the most important states and focus a disproportionate percentage of their resources on early states.

Thus we should expect that different candidates will follow similar resource allocation strategies as one another regardless of the order of the states. But if different candidates are allocating similar levels of resources to the various states, on average one would expect candidates’ decisions about resource allocation to have little net effect on the total fraction of voters in a state who prefer one candidate to the other. Thus the distribution of voter preferences in each state should not be significantly affected by allowing for endogenous candidate strategies.

---

9There is empirical evidence that candidates follow similar resource allocation strategies in presidential primaries. For instance, Brams and Davis (1982) note that presidential candidates in a given primary spent similar levels of financial resources as one another in each state in the 1976 presidential primaries.
The proposition we have given is formally proven by showing that there is an equilibrium in which candidates use the same resource allocation in each state. This result indicates that there are natural conditions under which allowing for endogenous candidate strategies will not affect the optimal order of the states given in the previous section. Since the distribution of voter preferences in each state remains the same after allowing for endogenous candidate strategies, the optimal order of the states also remains the same.

5. Conclusion

New Hampshire and Iowa’s position as the first primary and caucus in the presidential election cycle clearly has an impact on which candidate gains the nomination, and their position as small and not necessarily representative states has been sharply questioned. Moreover, officials in larger states often express displeasure at the relatively minor influence they hold by voting later in the process. For instance the acting Governor of New Jersey in 2005 put it thus: “No longer will New Jersey be an afterthought in selection of a candidate for our nation’s highest office [...] No longer will candidates just court our wallets; now they will court our votes.” (Chen, 2005). Indeed, perhaps the strongest evidence of dissatisfaction is the constant race to be the earliest primary, as in the 2008 primary season.

We have offered a model of presidential primaries which allows for bandwagon effects. We provided conditions under which it is, in fact, optimal for states like Iowa and New Hampshire to vote first, despite the fact that they may be “unrepresentative” compared with other more populous states. Allowing small states to vote first has the benefit of enabling larger states to observe the actions of smaller states and possibly make more accurate decisions as a result. And Iowa and New Hampshire may also receive more accurate signals of candidate abilities since their small size affords closer contact with the candidates and their residents have higher levels of education. This has the further benefit that there will be a greater chance that early states will successfully reveal the identity of the superior candidate before later states have to act.
Our results about the nature of optimal primaries are surprisingly robust. They extend to some settings where bad bandwagons might occur and to settings where the candidates may vary their allocations of campaign resources with the order of the states. Nonetheless, a limitation of our analysis is that we have only established the robustness result with the possibility of bad bandwagons in a two-state setting. It would be desirable to extend this to an arbitrary number of states—a topic we leave to future work.

APPENDIX

Proof of Proposition 1:

First note that if \( p \pi_N \geq \frac{1}{2} \), then the majority of voters vote for the better candidate in each state, the better candidate wins with probability 1, and the result holds trivially. Thus we assume that \( p \pi_N < \frac{1}{2} \) throughout this proof.

Suppose by means of contradiction that the ordering of the states that maximizes the probability that the higher quality candidate is elected is some ordering \( S(1), \ldots, S(m) \) distinct from the ordering described in the proposition. Then there are some states \( S(j) \) and \( S(j+1) \) in this ordering such that \( \lambda(j+1) < \lambda(j) \). We seek to show that the higher quality candidate would be elected with greater probability if the order of states \( S(j) \) and \( S(j+1) \) were reversed.

To do this, we show that if the state of the world is \( a \), then reversing the order of states \( S(j) \) and \( S(j+1) \) would increase the probability that candidate \( A \) is elected. A virtually identical argument shows that if the state of the world is \( b \), then reversing the order of states \( S(j) \) and \( S(j+1) \) would increase the probability that candidate \( B \) is elected. Letting \( \pi_A^{(j)} \) denote the fraction of \( A \)-partisans in the \( j^{th} \) state to vote, we consider three cases:

Case 1: Suppose that the values of \( \pi_A^{(k)} \) for \( k = 1, \ldots, j - 1 \) are such that there is some state \( S(k) \) with \( k < j \) for which the fraction of voters in state \( S(k) \) that votes for candidate \( A \) is greater than \( 1 - p \pi_N \). Then all neutral voters in states \( S(k) \) with \( k \geq j \) vote the same way regardless of the order of states \( S(j) \) and \( S(j+1) \), and reversing the order of these two states has no effect on the probability with which candidate \( A \) is elected.
Case 2: Suppose that the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j+1$ are such that there is no state $S(k)$ with $k \leq j+1$ for which the fraction of voters in state $S(k)$ that votes for candidate $A$ is greater than $1 - p\pi_N$. Then all neutral voters in states $S(k)$ with $k \geq j$ again vote the same way regardless of the order of states $S(j)$ and $S(j+1)$, and reversing the order of these two states again has no effect on the probability with which candidate $A$ is elected.

Case 3: If neither of the above possibilities holds, then the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j+1$ are such there is no state $S(k)$ with $k < j$ for which the fraction of voters in state $S(k)$ that votes for candidate $A$ is greater than $1 - p\pi_N$, and there is some state $S(k)$ with $k = j$ or $k = j+1$ for which the fraction of voters in state $S(k)$ that votes for candidate $A$ is greater than $1 - p\pi_N$. We show that, conditional on the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j+1$ satisfying these conditions, the probability that $A$ wins the election is strictly greater if the order of the states $S(j)$ and $S(j+1)$ is reversed.

If the values of $\pi_A^{(k)}$ satisfy these conditions, then either the fraction of voters in state $S(j)$ that votes for candidate $A$ is greater than $1 - p\pi_N$ or the fraction of voters in state $S(j)$ that votes for candidate $A$ is less than or equal to $1 - p\pi_N$ and the fraction of voters in state $S(j+1)$ that votes for candidate $A$ is greater than $1 - p\pi_N$. Conditional on the fraction of voters in state $S(j)$ that votes for candidate $A$ being greater than $1 - p\pi_N$, the fraction of voters in state $S(j)$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ and the fraction of voters in state $S(j+1)$ that votes for candidate $A$ is drawn from the uniform distribution on $[\pi_N, 1]$. And conditional on the fraction of voters in state $S(j)$ that votes for candidate $A$ being less than or equal to $1 - p\pi_N$ and the fraction of voters in state $S(j+1)$ that votes for candidate $A$ being greater than $1 - p\pi_N$, the fraction of voters in state $S(j)$ that votes for candidate $A$ is drawn from the uniform distribution on $[p\pi_N, 1 - p\pi_N]$ and the fraction of voters in state $S(j+1)$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$.

In particular, if the values of $\pi_A^{(k)}$ satisfy these conditions, then with probability $\frac{1 - \pi_N}{2(1 + 2p)\pi_N}$, the fraction of voters in state $S(j)$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ and the fraction of voters in state $S(j+1)$ that votes
for candidate A is drawn from the uniform distribution on \([\pi_N, 1]\). And with probability 
\[
\frac{1-2p\pi_N}{2-(1+2p)\pi_N},
\]
the fraction of voters in state \(S_{(j)}\) that votes for candidate A is drawn from the uniform distribution on \([p\pi_N, 1-p\pi_N]\) and the fraction of voters in state \(S_{(j+1)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\).

Now note that if \(\pi_N \geq 1 - p\pi_N\), then the uniform distribution on \([\pi_N, 1]\) strictly first order stochastically dominates the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\), and the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\) strictly first order stochastically dominates the uniform distribution on \([p\pi_N, 1-p\pi_N]\). Thus if \(\pi_N \geq 1 - p\pi_N\) and we reverse the order of the states \(S_{(j)}\) and \(S_{(j+1)}\) by allowing the smaller state to vote before the larger state, then the resulting distribution of the total fraction of voters in states \(S_{(j)}\) and \(S_{(j+1)}\) that votes for A strictly first order stochastically dominates the original distribution of the fraction of voters in states \(S_{(j)}\) and \(S_{(j+1)}\) that votes for A.

Now suppose that \(\pi_N < 1 - p\pi_N\). Rewriting the distribution of the fraction of voters in state \(S_{(j)}\) that votes for candidate A, we see that if the values of \(\pi_A^{(k)}\) satisfy the conditions in Case 3, then with probability \(\frac{1-(1+p)\pi_N}{2-(1+2p)\pi_N}\), the fraction of voters in state \(S_{(j)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\) and the fraction of voters in state \(S_{(j+1)}\) that votes for candidate A is drawn from the uniform distribution on \([\pi_N, 1-p\pi_N]\). With probability \(\frac{p\pi_N}{2-(1+2p)\pi_N}\), the fraction of voters in state \(S_{(j)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\) and the fraction of voters in state \(S_{(j+1)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1]\). With probability \(\frac{(1-p)\pi_N}{2-(1+2p)\pi_N}\), the fraction of voters in state \(S_{(j)}\) that votes for candidate A is drawn from the uniform distribution on \([p\pi_N, \pi_N]\) and the fraction of voters in state \(S_{(j+1)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\). And with probability \(\frac{1-(1+p)\pi_N}{2-(1+2p)\pi_N}\), the fraction of voters in state \(S_{(j)}\) that votes for candidate A is drawn from the uniform distribution on \([\pi_N, 1-p\pi_N]\) and the fraction of voters in state \(S_{(j+1)}\) that votes for candidate A is drawn from the uniform distribution on \([1-p\pi_N, 1-(1-p)\pi_N]\).
Note that it is equally likely that the fraction of voters in state $S_j$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ and the fraction of voters in state $S_{j+1}$ that votes for candidate $A$ is drawn from the uniform distribution on $[\pi_N, 1 - p\pi_N]$ as it is that the fraction of voters in state $S_j$ that votes for candidate $A$ is drawn from the uniform distribution on $[\pi_N, 1 - p\pi_N]$ and the fraction of voters in state $S_{j+1}$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$.

Thus the order of states $S_j$ and $S_{j+1}$ cannot affect the overall distribution of the total fraction of voters that votes for candidate $A$ in states $S_j$ and $S_{j+1}$ that arises from these circumstances.

Thus the only two ways the order of the states can matter is if the fraction of voters in state $S_j$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ and the fraction of voters in state $S_{j+1}$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ or if the fraction of voters in state $S_j$ that votes for candidate $A$ is drawn from the uniform distribution on $[p\pi_N, \pi_N]$ and the fraction of voters in state $S_{j+1}$ that votes for candidate $A$ is drawn from the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$.

Now the uniform distribution on $[1 - p\pi_N, 1]$ strictly first order stochastically dominates the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$. And the uniform distribution on $[1 - p\pi_N, 1 - (1 - p)\pi_N]$ strictly first order stochastically dominates the uniform distribution on $[p\pi_N, \pi_N]$. Thus in either of the two circumstances in the previous paragraph, if we reverse the order of the states $S_j$ and $S_{j+1}$ by allowing the smaller state to vote before the larger state, then the resulting distribution of the total fraction of voters in states $S_j$ and $S_{j+1}$ that votes for $A$ strictly first order stochastically dominates the original distribution of the fraction of voters in states $S_j$ and $S_{j+1}$ that votes for $A$.

Thus regardless of whether $\pi_N > 1 - p\pi_N$, if we make this change to the order in which states $S_j$ and $S_{j+1}$ vote, then the distribution of the total fraction of voters in the population that votes for candidate $A$ strictly first order stochastically dominates the distribution of the total fraction of voters in the population that votes for candidate $A$ under the original
order. From this it follows that reversing the order in which states $S_j$ and $S_{j+1}$ vote increases the probability that candidate $A$ is elected. □

**Proof of Proposition 2:**

First note that if $p_j \pi_N \geq \frac{1}{2}$ for the state $S_j$ with the largest value of $p_j$ and this state votes first, then the majority of voters vote for the better candidate in the first state, voters in later states learn which candidate is the better candidate from the results of the first state, the majority of voters vote for the better candidate in all future states, and the better candidate wins with probability 1. Thus if $p_j \pi_N \geq \frac{1}{2}$ for the state $S_j$ with the largest value of $p_j$, then the better candidate wins with probability 1 if the states are ordered as stated in the proposition and the result holds. Thus we assume that $p_j \pi_N < \frac{1}{2}$ for all states $S_j$ throughout this proof.

Suppose by means of contradiction that the ordering of the states that maximizes the probability that the higher quality candidate is elected is some ordering $S(1), \ldots, S(m)$ distinct from the ordering described in the proposition. Then there are some states $S_j$ and $S_{j+1}$ in this ordering such that $p_{j+1} > p_j$. We seek to show that the higher quality candidate would be elected with greater probability if the order of states $S_j$ and $S_{j+1}$ were reversed.

To do this, we show that if the state of the world is $a$, then reversing the order of states $S_j$ and $S_{j+1}$ would increase the probability that candidate $A$ is elected. A virtually identical argument shows that if the state of the world is $b$, then reversing the order of states $S_j$ and $S_{j+1}$ would increase the probability that candidate $B$ is elected. Letting $\pi_A^{(j)}$ denote the fraction of $A$-partisans in the $j^{th}$ state to vote, we consider two cases:

**Case 1:** Suppose that the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j - 1$ are such that there is some state $S_{(k)}$ with $k < j$ for which the fraction of voters in state $S_{(k)}$ that votes for candidate $A$ is greater than $1 - p_{(k)} \pi_N$. Then all neutral voters in states $S_{(k)}$ with $k \geq j$ vote the same way regardless of the order of states $S_j$ and $S_{j+1}$, and reversing the order of these two states has no effect on the probability with which candidate $A$ is elected.
Case 2: Suppose that the above possibility does not hold. Then the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j - 1$ are such that there is no state $S_{(k)}$ with $k < j$ for which the fraction of voters in state $S_{(k)}$ that votes for candidate $A$ is greater than $1 - p_{(k)}\pi_N$. We show that, conditional on the values of $\pi_A^{(k)}$ for $k = 1, \ldots, j - 1$ satisfying these conditions, the probability that $A$ wins the election is strictly greater if the order of states $S_{(j)}$ and $S_{(j+1)}$ is reversed.

Note that reversing the order of states $S_{(j)}$ and $S_{(j+1)}$ never affects the distribution of votes in the states $S_{(k)}$ with $k > j + 1$, so to illustrate that reversing the order of states $S_{(j)}$ and $S_{(j+1)}$ increases the probability of electing candidate $A$, it suffices to show that the distribution of the total fraction of voters in states $S_{(j)}$ and $S_{(j+1)}$ that votes for candidate $A$ when the order of these states is reversed strictly first order stochastically dominates the distribution of the total fraction of voters in these states that votes for candidate $A$ under the original order.

Note that if voters in the $j+1^{th}$ state to vote never voted differently as a result of how voters in the $j^{th}$ state to vote voted, then reversing the order of states $S_{(j)}$ and $S_{(j+1)}$ would not affect the distribution of the total fraction of voters in these states that votes for candidate $A$. The fraction of voters that votes for candidate $A$ in one of these states would be drawn from the uniform distribution on $[p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]$ and the fraction of voters that votes for candidate $A$ in the other state would be drawn from the uniform distribution on $[p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]$. We call the condition in which voters in the $j + 1^{th}$ state to vote never vote differently as a result of how voters in the $j^{th}$ state to vote voted condition $(\ast)$. Since the distribution of the total fraction of voters in states $S_{(j)}$ and $S_{(j+1)}$ that votes for candidate $A$ is the same under condition $(\ast)$ regardless of the order of these states, to illustrate how reversing the order of states $S_{(j)}$ and $S_{(j+1)}$ affects the distribution of the total fraction of voters in these states that votes for candidate $A$, it suffices to illustrate how removing condition $(\ast)$ would affect the distribution of the total fraction of voters in these states that votes for candidate $A$ under the two possible orderings of these states.
Note that if condition (*) is satisfied and the order of the states is not reversed, then with probability 
\[
\frac{2p_{(j+1)}-1)}{1-\pi_N},
\]
the fraction of voters that votes for A in the \(j\)th state to vote is drawn from the uniform distribution on \([1 - p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]\) and the fraction of voters that votes for A in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]\). And with probability \(\frac{1-2p_{(j)}\pi_N}{1-\pi_N}\), the fraction of voters that votes for A in the \(j\)th state to vote is drawn from the uniform distribution on \([p_{(j)}\pi_N, 1 - p_{(j)}\pi_N]\) and the fraction of voters that votes for A in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]\).

And if condition (*) is satisfied and the order of the states is reversed, then with probability 
\[
\frac{2p_{(j+1)}-1)}{1-\pi_N},
\]
the fraction of voters that votes for A in the \(j\)th state to vote is drawn from the uniform distribution on \([1 - p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]\) and the fraction of voters that votes for A in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]\). And with probability \(\frac{1-2p_{(j+1)}\pi_N}{1-\pi_N}\), the fraction of voters that votes for A in the \(j\)th state to vote is drawn from the uniform distribution on \([p_{(j+1)}\pi_N, 1 - p_{(j+1)}\pi_N]\) and the fraction of voters that votes for A in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]\).

If the order of the states is not reversed and the fraction of voters that would vote for A under condition (*) in the \(j\)th state to vote is drawn from the uniform distribution on \([1 - p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]\) and the fraction of voters that would vote for A under condition (*) in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]\), then removing condition (*) increases the total number of voters that votes for candidate A by \(\lambda(\pi_N - p_{(j+1)}\pi_N)\). Otherwise removing condition (*) has no effect on the number of voters that votes for candidate A.

And if the order of the states is reversed and the fraction of voters that would vote for A under condition (*) in the \(j\)th state to vote is drawn from the uniform distribution on \([1 - p_{(j+1)}\pi_N, 1 - (1 - p_{(j+1)})\pi_N]\) and the fraction of voters that would vote for A under condition (*) in the \(j + 1\)th state to vote is drawn from the uniform distribution on \([p_{(j)}\pi_N, 1 - (1 - p_{(j)})\pi_N]\), then removing condition (*) increases the total number of voters that votes for candidate A.
candidate $A$ by $\lambda(\pi_N - p(j)\pi_N)$, a greater amount than the corresponding increase under the original order of the states. Otherwise removing condition (*) has no effect on the number of voters that votes for candidate $A$.

Now suppose that the total number of voters that votes for candidate $A$ in states $S(j)$ and $S(j+1)$ under condition (*) is $x$. We seek to show that the probability the total number of voters that votes for candidate $A$ increases by removing condition (*) is at least as large when the order of these states is reversed as it is under the original order. We let $q(x)$ denote this probability conditional on $x$ under the original order of states $S(j)$ and $S(j+1)$ and let $r(x)$ denote this probability conditional on $x$ when the order of these states is reversed, and seek to show that $r(x) \geq q(x)$.

To see this, suppose that if the total number of voters in states $S(j)$ and $S(j+1)$ that votes for $A$ under condition (*) is $x$, then there is a positive probability that the total number of voters that votes for candidate $A$ will increase by removing condition (*) under the original order of states $S(j)$ and $S(j+1)$. In this case, $x = \lambda(y + z)$ for some $y \in [1 - p(j)\pi_N, 1 - (1 - p(j))\pi_N]$ and $z \in [p(j+1)\pi_N, 1 - (1 - p(j+1))\pi_N]$.

Now if $v(y) = y - p(j)\pi_N + p(j+1)\pi_N$ and $w(z) = z + p(j)\pi_N - p(j+1)\pi_N$, then $v(y)$ and $w(z)$ are fractions satisfying $v(y) \in [1 - p(j+1)\pi_N, 1 - (1 - p(j+1))\pi_N]$, $w(z) \in [p(j)\pi_N, 1 - (1 - p(j))\pi_N]$, and $x = \lambda(v(y) + w(z))$. From this it follows that the measure of the set of values of $y \in [1 - p(j)\pi_N, 1 - (1 - p(j))\pi_N]$ for which $x = \lambda(y + z)$ for some $z \in [p(j+1)\pi_N, 1 - (1 - p(j+1))\pi_N]$ is no greater than the measure of the set of values of $v \in [1 - p(j+1)\pi_N, 1 - (1 - p(j+1))\pi_N]$ for which $x = \lambda(v + w)$ for some $w \in [p(j)\pi_N, 1 - (1 - p(j))\pi_N]$ since if $Y$ denotes the set of $y \in [1 - p(j)\pi_N, 1 - (1 - p(j))\pi_N]$ for which $x = \lambda(y + z)$ for some $z \in [p(j+1)\pi_N, 1 - (1 - p(j+1))\pi_N]$, then the measure of $Y$ is equal to the measure of $v(Y)$. Thus, conditional on the total number of voters in states $S(j)$ and $S(j+1)$ that votes for $A$ under condition (*) being $x$, there are at least as many circumstances under which the total number of voters that votes for candidate $A$ is increased by removing condition (*) when the order of states $S(j)$ and $S(j+1)$ is reversed as there are circumstances under which this total is increased under the original order of these states.
But this means that for any \( x \) such that there is a positive probability that the total number of voters that votes for candidate \( A \) is increased by removing condition (*) under the original order of states \( S_{(j)} \) and \( S_{(j+1)} \), we have \( r(x) \geq q(x) \). Furthermore, in the cases in which the total number of voters that votes for candidate \( A \) is increased by removing condition (*), this increase is greater when the order of states \( S_{(j)} \) and \( S_{(j+1)} \) is reversed than it is under the original order of these states. This indicates that the distribution of the total fraction of voters in states \( S_{(j)} \) and \( S_{(j+1)} \) that votes for candidate \( A \) when the order of these states is reversed strictly first order stochastically dominates the distribution of the total fraction of voters in these states that votes for candidate \( A \) under the original order.

But this means that if we make this change to the order in which states \( S_{(j)} \) and \( S_{(j+1)} \) vote, then the distribution of the total fraction of voters in the population that votes for candidate \( A \) strictly first order stochastically dominates the distribution of the total fraction of voters in the population that votes for candidate \( A \) under the original order. From this it follows that reversing the order in which states \( S_{(j)} \) and \( S_{(j+1)} \) vote increases the probability that candidate \( A \) is elected. \( \square \)

**Proof of Proposition 3:**

Since \( \delta < \max\{(2p-1)\pi_N, \frac{1-2p\pi_N}{2}\} \), it follows that if \( y(1) - p\pi_N \in [0, \delta] \), then \( y(1) - (1-p)\pi_N \in [\delta, 1 - \pi_N - \delta] \) and if \( y(1) - (1-p)\pi_N \in [1 - \pi_N - \delta, 1 - \pi_N] \), then \( y(1) - p\pi_N \in [\delta, 1 - \pi_N - \delta] \).

From the assumptions on \( f(x) \) we know that \( \pi_{(2)} < 1 - p \) if and only if \( y(1) - p\pi_N < \delta \) and \( y(1) - (1-p)\pi_N \in [\delta, 1 - \pi_N - \delta] \) or \( y(1) - p\pi_N < 0 \). Combining this with the first result in this proof indicates that \( \pi_{(2)} < 1 - p \) if and only if \( y(1) - p\pi_N < \delta \) or \( y(1) < \delta + p\pi_N \).

Similarly, from the assumptions on \( f(x) \) we know that \( \pi_{(2)} > p \) if and only if either \( y(1) - (1-p)\pi_N > 1 - \pi_N - \delta \) and \( y(1) - p\pi_N \in [\delta, 1 - \pi_N - \delta] \) or \( y(1) - (1-p)\pi_N > 1 - \pi_N \). Combining this with the first result in this proof indicates that \( \pi_{(2)} > p \) if and only if \( y(1) - (1-p)\pi_N > 1 - \pi_N - \delta \) or \( y(1) > 1 - \delta - p\pi_N \).

Now suppose the state of the world is \( a \). In this case \( \pi_A^{(1)} = y(1) - p\pi_N \). Combining this with the results in the previous two paragraphs shows that \( \pi_{(2)} < 1 - p \) if and only if \( \pi_A^{(1)} < \delta \)
and $\pi_{(2)} > p$ if and only if $\pi_A^{(1)} > 1 - \delta - 2p\pi_N$. Thus all neutral voters in the second state vote for $B$ if and only if $\pi_A^{(1)} < \delta$, all neutral voters in the second state vote for $A$ if and only if $\pi_A^{(1)} > 1 - \delta - 2p\pi_N$, and all neutral voters vote according to their private signal if and only if $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N$.

Note that if $\pi_A^{(1)} < \delta$ and all neutral voters in the second state vote for $B$, then the majority of voters in both states vote for $B$, and $B$ is elected regardless of the order of the states. And if $\pi_A^{(1)} > 1 - \delta - 2p\pi_N$ and all neutral voters in the second state vote for $A$, then the majority of voters in both states vote for $A$, and $A$ is elected regardless of the order of the states. Thus the only way the order of the states can affect which candidate is elected is if $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N$.

Now if $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N$, then the total number of voters who vote for $A$ is $\lambda(1)(\pi_A^{(1)} + p\pi_N) + \lambda(2)(\pi_A^{(2)} + p\pi_N)$. Thus candidate $A$ wins the election if and only if $\lambda(1)(\pi_A^{(1)} + p\pi_N) + \lambda(2)(\pi_A^{(2)} + p\pi_N) \geq \frac{1}{2} (\lambda(1) + \lambda(2))$ or $\lambda(1)(\pi_A^{(1)} + p\pi_N - \frac{1}{2}) + \lambda(2)(\pi_A^{(2)} + p\pi_N - \frac{1}{2}) \geq 0$ or $\frac{\lambda(1)}{\lambda(2)}(\pi_A^{(1)} + p\pi_N - \frac{1}{2}) + \pi_A^{(2)} + p\pi_N - \frac{1}{2} \geq 0$ or $\pi_A^{(2)} \geq \frac{1}{2} - p\pi_N - \frac{\lambda(1)}{\lambda(2)}(\pi_A^{(1)} + p\pi_N - \frac{1}{2})$. So for a fixed $\pi_A^{(1)}$ satisfying $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N$, candidate $A$ wins the election with probability $1 - F\left(\frac{1}{2} - p\pi_N - \frac{\lambda(1)}{\lambda(2)}(\pi_A^{(1)} + p\pi_N - \frac{1}{2})\right)$, where $I$ let $F$ denote the cumulative distribution function satisfying $F = F_1 = F_2$.

From this it follows that if $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N$, then candidate $A$ wins the election with probability $\int_{\delta}^{1-\delta-2p\pi_N} \left[1 - F\left(\frac{1}{2} - p\pi_N - \frac{\lambda(1)}{\lambda(2)}(\pi_A^{(1)} + p\pi_N - \frac{1}{2})\right)\right] f(\pi_A^{(1)}) \, d\pi_A^{(1)} / Pr(\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N) = \int_{\delta}^{1-\delta-2p\pi_N} \left[1 - F\left(\frac{1}{2} - p\pi_N - \frac{\lambda(1)}{\lambda(2)}x\right)\right] f\left(\frac{1}{2} - p\pi_N + x\right) \, dx / Pr(\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2p\pi_N)$. Now assume without loss of generality that $\lambda_1 = \min\{\lambda_1, \lambda_2\}$ and $\lambda_2 = \max\{\lambda_1, \lambda_2\}$. These expressions indicate that candidate $A$ wins with greater probability when the small state votes first and only if $\int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x\right) f\left(\frac{1}{2} - p\pi_N + x\right) \, dx \leq \int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x\right) f\left(\frac{1}{2} - p\pi_N + x\right) \, dx$. And since $f(x)$ is constant for all $x \in [\delta, 1 - p\pi - \delta]$, it follows that candidate $A$ wins with greater probability when the small state votes first if and only if $\int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x\right) \, dx \leq \int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x\right) \, dx$.

Now $\int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x\right) \, dx = \int_{0}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x\right) + F\left(\frac{1}{2} - p\pi_N + \frac{\lambda_1}{\lambda_2} x\right) \, dx$ and $\int_{\delta + p\pi_N - \frac{1}{2}}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x\right) \, dx = \int_{0}^{1-\delta-2p\pi_N} F\left(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x\right) + F\left(\frac{1}{2} - p\pi_N + \frac{\lambda_2}{\lambda_1} x\right) \, dx$. 

26
And $\frac{d}{dx}[F(\frac{1}{2} - p\pi_N - u) + F(\frac{1}{2} - p\pi_N + u)] = [f(\frac{1}{2} - p\pi_N + u) - f(\frac{1}{2} - p\pi_N - u)] \geq 0$ since $f(x)$ is symmetric and weakly single-peaked about $x = \frac{1 - p\pi_N}{2} \geq \frac{1}{2} - p\pi_N$. Thus since $\lambda_2 > \lambda_1$, $F(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x) + F(\frac{1}{2} - p\pi_N + \frac{\lambda_1}{\lambda_2} x)$ is at least as large as $F(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x) + F(\frac{1}{2} - p\pi_N + \frac{\lambda_1}{\lambda_2} x)$ for any given $x > 0$. From this it follows that $\int_{0}^{\frac{1}{2} - \delta - \frac{\lambda_2}{\lambda_1} \pi_N} F(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x) + F(\frac{1}{2} - p\pi_N + \frac{\lambda_2}{\lambda_1} x) \, dx \leq \int_{0}^{\frac{1}{2} - \delta - \frac{\lambda_1}{\lambda_2} \pi_N} F(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x) + F(\frac{1}{2} - p\pi_N + \frac{\lambda_1}{\lambda_2} x) \, dx$ and $\int_{\delta + \frac{\lambda_2}{\lambda_1} \pi_N}^{\frac{1}{2}} F(\frac{1}{2} - p\pi_N - \frac{\lambda_1}{\lambda_2} x) \, dx \leq \int_{\delta + \frac{\lambda_2}{\lambda_1} \pi_N}^{\frac{1}{2}} F(\frac{1}{2} - p\pi_N - \frac{\lambda_2}{\lambda_1} x) \, dx$.

Thus candidate $A$ wins with greater probability when the small state votes first. And a similar argument shows that if the state of the world is $b$, then candidate $B$ wins with greater probability when the small state votes first. From this it follows that the better candidate wins with greater probability when the smaller state votes first. \(\Box\)

**Proof of Proposition 4:**

Since $\delta < \max\{(2p_j - 1)\pi_N, \frac{1 - 2p_j \pi_N}{2}\}$ for $j = 1$ and $2$, it follows that if $y_j - p_j \pi_N \in [0, \delta]$, then $y_j - (1 - p_j) \pi_N \in [\delta, 1 - \pi_N - \delta]$ and if $y_j - (1 - p_j) \pi_N \in [1 - \pi_N - \delta, 1 - \pi_N]$, then $y_j - p_j \pi_N \in [\delta, 1 - \pi_N - \delta]$. From the assumptions on $f(x)$ we know that $\pi_{(2)} < 1 - p_{(2)}$ if and only if $y_{(1)} - p_{(1)} \pi_N < \delta$ and $y_{(1)} - (1 - p_{(1)}) \pi_N \in [\delta, 1 - \pi_N - \delta]$ or $y_{(1)} - p_{(1)} \pi_N < 0$. Combining this with the first result in this proof indicates that $\pi_{(2)} < 1 - p_{(2)}$ if and only if $y_{(1)} - p_{(1)} \pi_N < \delta$ or $y_{(1)} < \delta + p_{(1)} \pi_N$.

Similarly, from the assumptions on $f(x)$ we know that $\pi_{(2)} > p_{(2)}$ if and only if either $y_{(1)} - (1 - p_{(1)}) \pi_N > 1 - \pi_N - \delta$ and $y_{(1)} - p_{(1)} \pi_N \in [\delta, 1 - \pi_N - \delta]$ or $y_{(1)} - (1 - p_{(1)}) \pi_N > 1 - \pi_N$. Combining this with the first result in this proof indicates that $\pi_{(2)} > p_{(2)}$ if and only if $y_{(1)} - (1 - p_{(1)}) \pi_N > 1 - \pi_N - \delta$ or $y_{(1)} > 1 - \delta - p_{(1)} \pi_N$.

Now suppose that the state of the world is $a$. In this case $\pi_{(1)}^a = y_{(1)} - p_{(1)} \pi_N$. Combining this with the results in the previous two paragraphs shows that $\pi_{(2)} < 1 - p_{(2)}$ if and only if $\pi_{(1)}^a < \delta$ and $\pi_{(2)} > p_{(2)}$ if and only if $\pi_{(1)}^a > 1 - \delta - 2p_{(1)} \pi_N$. Thus all neutral voters in the second state vote for $B$ if and only if $\pi_{(1)}^a < \delta$, all neutral voters in the second state vote for $A$ if and only if $\pi_{(1)}^a > 1 - \delta - 2p_{(1)} \pi_N$, and all neutral voters vote according to their private signal if and only if $\delta \leq \pi_{(1)}^a \leq 1 - \delta - 2p_{(1)} \pi_N$. \(27\)
From this it follows that if $\pi_A^{(1)} < \delta$, then the total number of voters who vote for $A$ is $\lambda(\pi_A^{(1)} + p(1)\pi_N + \pi_A^{(2)})$. If $\delta \leq \pi_A^{(1)} \leq 1 - \delta - 2\max\{p(1), p(2)\}\pi_N$, then the total number of voters who vote for $A$ is $\lambda(\pi_A^{(1)} + p(1)\pi_N + \pi_A^{(2)} + p(2)\pi_N)$. If $1 - \delta - 2\max\{p(1), p(2)\}\pi_N < \pi_A^{(1)} \leq 1 - \delta - 2\min\{p(1), p(2)\}\pi_N$, then the total number of voters who vote for $A$ is $\lambda(\pi_A^{(1)} + p(1)\pi_N + \pi_A^{(2)} + p(2)\pi_N)$ if $p(1) < p(2)$ and $\lambda(\pi_A^{(1)} + p(1)\pi_N + \pi_A^{(2)} + \pi_N)$ otherwise. And if $\pi_A^{(1)} > 1 - \delta - 2\min\{p(1), p(2)\}\pi_N$, then the total number of voters who vote for $A$ is $\lambda(\pi_A^{(1)} + p(1)\pi_N + \pi_A^{(2)} + \pi_N)$.

But in any of these cases, the total number of voters who vote for $A$ is at least as large if the state with the more accurate private signal votes first. Thus the probability that $A$ is elected is maximized when the state with the more accurate private signal votes first. A similar argument shows that if the state of the world is $b$, then the probability that $B$ is elected is maximized when the state with the more accurate private signal votes first. Thus the probability that the better candidate is elected is maximized when the state with the more accurate private signal votes first. □

Proof of Proposition 5:

Note that the candidate budget allocation game is a strictly competitive game because a candidate can only increase his or her probability of winning the election by decreasing the other candidate’s probability of winning the election. Thus by von Neumann (1928), we know that strategies for the candidates are equilibrium strategies if and only if each candidate is using a minimax strategy or a strategy which minimizes the maximum payoff the opposing candidate can obtain against the original candidate’s strategy.

From this it follows that if there is a pure strategy equilibrium in which candidate $A$ uses the allocation $a^*$, then $a^*$ is a minimax strategy for candidate $A$. Thus if $Pr(A \text{ wins}|(\sigma_A, \sigma_B))$ denotes the probability with which candidate $A$ believes he will win the election when candidate $A$ uses the strategy $\sigma_A$ and candidate $B$ uses the strategy $\sigma_B$, then $\sigma_A = a^*$ maximizes the value of $\min_{\sigma_B} Pr(A \text{ wins}|(\sigma_A, \sigma_B))$. 28
Now if $Pr(B \text{ wins} | (\sigma_A, \sigma_B))$ denotes the probability with which candidate $B$ believes he will win the election when candidate $A$ uses the strategy $\sigma_A$ and candidate $B$ uses the strategy $\sigma_B$, then $Pr(B \text{ wins} | (\sigma_B, \sigma_A)) = Pr(A \text{ wins} | (\sigma_A, \sigma_B))$. To see this, note that it is equally likely that the fraction of $A$-partisans in state $S_j$ before the candidates campaign is some fraction $\pi_j$ as it is that the fraction of $B$-partisans in state $S_j$ before the candidates campaign is $\pi_j$. Thus it is equally likely that the fraction of $A$-partisans in state $S_j$ after the candidates campaign is $\pi_j'(\pi_j, a_j, b_j)$ if candidate $A$ chooses $a = (a_1, a_2, \ldots, a_m)$ and candidate $B$ chooses $b = (b_1, b_2, \ldots, b_m)$ as it is that the fraction of $B$-partisans in state $S_j$ after the candidates campaign is $\pi_j'(\pi_j, a_j, b_j)$ if candidate $B$ chooses $a = (a_1, a_2, \ldots, a_m)$ and candidate $A$ chooses $b = (b_1, b_2, \ldots, b_m)$. And since both candidates believe that each candidate is the higher quality candidate with probability $\frac{1}{2}$, each candidate believes it is equally likely that candidate $A$ will win the election if the fraction of $A$-partisans in each state $S_j$ is $\pi_j'(\pi_j, a_j, b_j)$ as it is that candidate $B$ will win the election if the fraction of $B$-partisans in each state $S_j$ is $\pi_j'(\pi_j, a_j, b_j)$. Putting this together shows that $Pr(B \text{ wins} | (\sigma_B, \sigma_A)) = Pr(A \text{ wins} | (\sigma_A, \sigma_B))$.

But this means that if $\sigma_A = a^*$ maximizes the value of $\min_{\sigma_B} Pr(A \text{ wins} | (\sigma_A, \sigma_B))$, then $\sigma_B = a^*$ also maximizes the value of $\min_{\sigma_B} Pr(B \text{ wins} | (\sigma_A, \sigma_B))$, and $\sigma_B = a^*$ is a minimax strategy for candidate $B$. Thus if $\sigma_A = a^*$ is a minimax strategy for candidate $A$, then it is a pure strategy equilibrium for the candidates to use the allocation in which they both choose $a^*$. But if both candidates use the allocation $a^*$, then the fraction of $A$-partisans in state $S_j$ is the same after candidates choose budget allocations than it is before the candidates choose budget allocations. Thus the distribution of voter preferences in each state is the same as in the original model without endogenous candidate strategies. □

References


