Journal of Public Economics 109 (2014) 64–75

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Optimal primaries

Patrick Hummel [a,*], Richard Holden [b]

[a] Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
[b] University of New South Wales, Room 470B, Australian School of Business, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 1 October 2012
Received in revised form 16 October 2013
Accepted 17 October 2013
Available online 31 October 2013

Keywords:
Optimal primaries
Order of states
Social planning

ABSTRACT

We analyze a model of US presidential primary elections for a given 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. 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 pre-determined 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.[1] 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 more likely to vote for candidates who were successful in early states. Theoretical models such as Ali and Kartik (2012) show that such behavior has a rational basis.[2] Moreover, there is significant empirical support that these momentum effects play an important role (Bartels, 1985, 1988; Kenney and Rice, 1994; Knight and Schiff, 2010; Popkin, 1994).[3]

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".[4] 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% of delegates are decided by March 2 today compared to 10% 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. Norrader (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.

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.

* Corresponding author.
E-mail addresses: phummel@google.com (P. Hummel), richard.holden@unsw.edu.au (R. Holden).

[1] The rules for primaries are set by the by-laws committees of the parties. The several states are expected to comply with these. They typically do, with some exceptions.

[2] In addition, classic papers on social learning such as Banerjee (1992) and Bikhchandani et al. (1992) have shown that information cascades can be rational in other economic contexts. For more work on sequential voting, see Battaglini (2005), Callander (2007), and Dekel and Piccione (2000).

[3] 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.

[4] As quoted in *The West Wing*.

0047-2727/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jpubeco.2013.10.001

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 as 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.

We derive conditions under which a social planner can maximize 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 where voters receive more accurate private information to vote earlier typically maximizes the probability that the higher quality candidate is elected. Finally, we discuss the robustness of the results when we explicitly model the possibility that candidate strategies are an endogenous function of the order in which the states vote.

The intuition behind the optimal ordering of the states is as follows. Voters in later states will be able to use the information about how early states voted to get a better sense of which candidate is the best candidate. Thus having larger states vote later enables a larger number of people to use the information from the results of early states, which in turn means that more people will be able to make a more well-informed decision when they vote. Similarly, allowing states with well-informed voters to vote earlier means that voters in less informed states will be able to use the information about how well-informed voters voted, which again means people will be able to make more well-informed decisions when they 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 better informed voters 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.[5]

Furthermore, when candidates campaign in small states, voters are more likely to be able to meet the candidates individually and obtain precise information 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.[6] And a few weeks before the 1996 New Hampshire Republican primary, approximately 20% 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.[7] Vavreck et al. (2002) provide 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."

This paper contributes to a literature on the optimal way to design a primary election. Other works on optimal primaries such as Callander (2007), Deltas et al. (2012), and Hummel and Knight (2012) have addressed questions related to whether a social planner should prefer simultaneous elections to sequential elections. However, there has been almost no work that addresses the question of the optimal way to order states in a presidential primary, given that this primary will use a sequential election. This is the subject of the current paper.

2. The model

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 λ_j and may allocate a total of λ_j delegates to the candidates. The candidate who receives the most delegates wins the election and ties are broken randomly. We consider two possible ways that states may allocate their delegates to the candidates.

The first method we consider is one in which states allocate their delegates to candidates in direct proportion to the number of votes each candidate received. Under this method, a candidate wins the election if and only if the candidate received a majority of votes. 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.

The second method we consider is a scenario in which a state allocates all of its delegates to the candidate who received the most votes in that state. We consider both of these possibilities because some states allocate all of their delegates to the candidate who received the most votes in that state, and other states allocate their delegates in close proportion to the number of votes each candidate received in the state.

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. In each state S_j, a fraction κ_j of the voters observe the state of the world directly, and the remaining voters do not. These remaining voters each share a common prior that the probability the state of the world is a is ϕ for some $\phi \in (0, 1)$.

If the state of the world is $a(b)$, then each imperfectly informed 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 $\alpha(\beta)$ with probability $p_j > \frac{1}{2}$ and the value $\beta(\alpha)$ with probability $1 - p_j$. We assume that $p_j > \max\{\phi, 1 - \phi\}$ for all j to ensure that a voter will initially believe it is more likely that candidate A is a better candidate after receiving the signal α and candidate B is a better candidate after receiving the signal β.

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 $A(B)$ in state $a(b)$, and utility 0 from electing candidate $B(A)$ in state $a(b)$.[8] Throughout we let π_j^A denote the fraction of voters in state S_j who are A-partisans, π_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

5 See, for instance, US Census Bureau, 1990 Census of Population, CPH-L-96 and US Census Bureau, 2000 Census of Population, P37.

6 This was noted in a November 2007 CBS News/New York Times Poll.

7 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 Vavreck et al. (2002) claim that New Hampshire's small size creates ideal opportunities for voters to interact with the candidates.

8 Our model is relevant in cases where a general election follows the primary election if some voters are partisans who are uninterested in candidate quality, while other neutral voters make their decisions, in part, based on the electability of the candidates.

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 by voting for the candidate they like best given their current information. 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 $\frac{1}{2}$.

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. This is a sensible objective for the social planner because, for example, if the majority of voters are neutrals, then maximizing the probability the higher quality candidate is elected also maximizes the welfare of the majority of the voters in the primary. While the planner and the voters know the values of π_j^N, we assume that the planner and the voters do not know the values of π_j^A and π_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 π_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.

This model differs from Selman (2010) in that the values of π_j^A and π_j^B in each state are not common knowledge in our model. If these values were common knowledge, then voters could deduce the identity of the superior candidate with virtual certainty after the election in the first state. By contrast, in our model, the uncertainty about the values of π_j^A and π_j^B means that there may be uncertainty about the identity of the superior candidate even after several states hold their elections.

2.1. 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 a few types of voters. The restriction to elections with two candidates is almost universal in formal theory literature on sequential voting and presidential primaries (Ali and Kartik, 2012; Battaglini, 2005; Battaglini et al., 2007; Brams and Davis, 1982; Callander, 2007; Dekel and Piccione, 2000; Klumpp and Polborn, 2006; Strumpf, 2002), and we do not believe that it is consequential for the results. In particular, the overall intuition for the results mentioned in the Introduction is an intuition that we would expect to continue to hold even when there are multiple candidates. Furthermore, 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.

It is also worth noting that 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.

We also note why we believe it is appropriate to consider uncertainty in the fraction of partisans rather than uncertainty in the fraction of neutrals. Had we considered an analogous model in which the relative fraction of A-partisans and B-partisans in each state were known, but the fraction of neutrals were uncertain, a voter would typically be able to infer which candidate is the higher quality candidate with certainty from the results of one state's election simply by noting whether the fraction of the vote A received in the state was greater or less than the known fraction of A-partisans in the state. However, incorporating uncertainty about the fraction of partisans makes the model more interesting because if a candidate does well, there can be uncertainty about whether the candidate did well because there were many partisan supporters for the candidate or if the candidate did well because the candidate was actually a higher quality candidate. Since uncertainty about the fraction of partisans enables voters to update their beliefs about the candidates in a realistic manner, but uncertainty about the fraction of neutrals does not, we focus on uncertainty in the fraction of partisans.

3. Results

Suppose that all neutral voters in state S_j who receive a private signal α vote for A and all neutral voters in state S_j who receive a private signal β 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 $\pi_j^A + (p_j(1-\kappa_j) + \kappa_j)\pi_j^N$, and if the state of the world is b, the fraction of voters that votes for candidate A in state S_j is $\pi_j^A + (1-p_j)(1-\kappa_j)\pi_j^N$.

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 $\pi_j^A = y_j - (p_j(1-\kappa_j) + \kappa_j)\pi_j^N$ or the state of the world is b and $\pi_j^A = y_j - (1-p_j)(1-\kappa_j)\pi_j^N$. From this it follows that if all neutral voters in states $S_{(1)}, \ldots, S_{(j-1)}$ vote according to their private signals and $\phi_{(j)}$ denotes the probability with which an imperfectly informed outside observer believes the state of the world is a after states $S_{(1)}, \ldots, S_{(j-1)}$ have voted, then

$$\phi_{(j)} = \frac{\phi_{(j-1)} f_{(j-1)}\left(w_{(j-1)}\right)}{\phi_{(j-1)} f_{(j-1)}\left(w_{(j-1)}\right) + \left(1-\phi_{(j-1)}\right) f_{(j-1)}\left(z_{(j-1)}\right)},$$

where $w_{(j-1)} \equiv y_{(j-1)} - (p_{(j-1)}(1-\kappa_{(j-1)}) + \kappa_{(j-1)})\pi_{(j-1)}^N$ and $z_{(j-1)} \equiv y_{(j-1)} - (1-p_{(j-1)})(1-\kappa_{(j-1)})\pi_{(j-1)}^N$.[9]

Now we specialize to the case in which $\pi_j^N = \pi^N$ for all j and π_j^A 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 $x \notin [0, 1-\pi^N]$. Under these assumptions, if all neutral voters in state $S_{(j-1)}$ vote according to their private signals and $y_{(j-1)} > 1 - (p_{(j-1)}(1-\kappa_{(j-1)}) + \kappa_{(j-1)})\pi^N$, then $\phi_{(j)} = 1$, as the only way so many voters could have voted for candidate A in state $S_{(j-1)}$ is if the state of the world is a. If $y_{(j-1)} < (p_{(j-1)}(1-\kappa_{(j-1)}) + \kappa_{(j-1)})\pi^N$, then $\phi_{(j)} = 0$, as the only way so few voters could have voted for candidate A in state $S_{(j-1)}$ is if the state of the world is b. But if $(p_{(j-1)}(1-\kappa_{(j-1)}) + \kappa_{(j-1)})\pi^N \le y_{(j-1)} \le 1 - (p_{(j-1)}(1-\kappa_{(j-1)}) + \kappa_{(j-1)})\pi^N$, then $f_{(j-1)}(w_{(j-1)}) = f_{(j-1)}(z_{(j-1)})$ and $\phi_{(j)} = \phi_{(j-1)}$.

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 $1 - (p_{(k)}(1-\kappa_{(k)}) + \kappa_{(k)})\pi^N$ in some previous state $S_{(k)}$, 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 this in some previous state $S_{(k)}$, 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 $\pi_j^N = \pi^N \ge \frac{1}{2}$ for all j, π_j^A is drawn from the uniform distribution on $[0, 1-\pi^N]$ for all j, $p_j = p > \frac{1}{2}$ for all j, and $\kappa_j = \kappa$ 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 λ_j, state $S_{(2)}$ has the second smallest value of λ_j, and in general state $S_{(k)}$ has the kth smallest value of λ_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

[9] This follows because the prior probability the state of the world is a is $\phi_{(j-1)}$ and the relative likelihood that A receives a fraction $y_{(j-1)}$ of the vote in state $S_{(j-1)}$ conditional the state of the world being a is $f(w_{(j-1)})$. Also, the prior probability the state of the world is b is $1-\phi_{(j-1)}$ and the relative likelihood that A receives a fraction $y_{(j-1)}$ of the vote in state $S_{(j-1)}$ conditional the state of the world being b is $f(z_{(j-1)})$.

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 in either case, at least as many voters vote for the higher quality candidate if the small states vote before the big states.

One can further illustrate how the optimal order of the states depends on the accuracy of the voters' private information in each state. To address this, we consider what happens when the states can be ordered in terms of how well-informed they are in the sense that states with more voters who observe the state of the world exactly (larger values of κ_j) also receive more informative private signals (larger values of p_j). Formally we assume that $p_j > p_k$ if and only if $\kappa_j > \kappa_k$ for all j and k. Under this assumption, we derive how the optimal order of the states depends on the level of informativeness of the states in Proposition 2:

Proposition 2. *Suppose that $\pi_j^N = \pi^N$ for all j, π_j^A is drawn from the uniform distribution on $[0, 1 - \pi^N]$ for all j, $\lambda_j = \lambda$ for all j, and $p_j > p_k$ if and only if $\kappa_j > \kappa_k$ for all j and k. 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 most informed to least informed, 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 kth 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 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, and there will be more later states where all neutral voters vote for the higher quality candidate if states with more accurate 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 the number of voters who vote for the higher quality candidate in both early states and later states. The result then follows.

Propositions 1 and 2 have both made use of the assumption that the fraction of partisans is drawn from a uniform distribution. This assumption is natural in many scenarios since it is quite common for individuals to only react to news if the news provides very strong evidence that an individual should change his or her beliefs, and such a reaction is captured by the uniform distribution.[10] It is also worth noting that this assumption is not completely necessary for these results because if the distribution of partisans is close to a uniform distribution, then the results in Propositions 1 and 2 still hold.

However, the assumption that the fraction of partisans is drawn from a distribution that is close to 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 δ 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 δ 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 δ, then imperfectly informed 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 imperfectly informed 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 otherwise. 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 δ satisfying $\delta < \max\left\{\left(\left(2p_j - 1\right)(1-\kappa_j) + \kappa_j\right)\pi^N, \frac{1}{2} - \left(p_j + \left(1 - p_j\right)\kappa_j\right)\pi^N\right\}$ for all j such that*

- *$f(x)$ is constant for $x \in [\delta, 1 - \pi^N - \delta]$,*
- *$\frac{\pi f(x)}{\pi f(x) + (1-\pi) f(z)} < \min\left\{1 - p_j\right\}$ if $x < \delta$ and $z \in [\delta, 1 - \pi^N - \delta]$, and*
- *$\frac{\pi f(x)}{\pi f(x) + (1-\pi) f(z)} > \max\left\{p_j\right\}$ if $x \in [\delta, 1 - \pi^N - \delta]$ and $z > 1 - \pi^N - \delta$.*

Under such a density $f(x)$, the fraction of A-partisans is considerably more likely to take on a particular value between δ and $1 - \pi^N - \delta$ than a particular value less than δ 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 δ, then imperfectly informed neutral voters in the second state all vote for the other candidate. But if both candidates do moderately well in the first state, then imperfectly informed voters continue to act on their private information in the second state. Under these assumptions we obtain this result:

Proposition 3. *Suppose that $m = 2$, $\pi_j^N = \pi^N$, $p_j = p$, $\kappa_j = \kappa$, and $f_j = f$ for all j, $(1-\kappa)\pi^N \geq \frac{1}{2}$, 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, so A wins the election. 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

[10] There is a long literature documenting that individuals underreact to news that contradicts their initial beliefs such as Frazzini (2006), Hirshleifer et al. (2009), and Ikenberry and Ramnath (2002).

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 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 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 $m=2$, $\pi_j^N=\pi^N$, $\lambda_j=\lambda$, $\kappa_j=\kappa$, 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 signals votes first and the state with less informative signals votes last.*

To understand why this holds, note that if the state with the more accurate 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 for similar reasons. Thus more favorable types of bandwagons occur when the state with the more accurate signals votes first.

At the same time, conditional on a bandwagon taking place, one would prefer that the state with the more accurate signals votes first. If a bandwagon takes place, then imperfectly informed 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 signals. Thus if a bandwagon takes place, more voters vote for the higher quality candidate if the state with the more accurate 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 signals vote first can only increase the number of voters that votes for the higher quality candidate.

Finally we address the question of the optimal order of the states when there are a large number of states and there is the possibility of bad bandwagons. First we address questions related to the optimal order of the states when the states may differ in their size but not in the precision of their private information. For simplicity here we focus on the situation in which a state allocates all of its delegates to the candidate who received the majority of the votes in that state and we also assume that $\kappa_j=0$ for all states j so there are no voters who observe the identity of the higher quality candidate with certainty. Under these assumptions, we prove the following result:

Proposition 5. *Suppose that $\pi_j^N=\pi^N$, $p_j=p$, $\kappa_j=0$, and $f_j=f$ for all j, $\pi^N\geq\frac{1}{2}$, Condition 1 holds, and the states allocate their delegates to the candidate who received the majority of votes in that state. 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.*

Thus when states allocate all of their delegates to the candidate who received the majority of votes in that state, it is optimal to order the states from smallest to largest even when there are a large number of states and there is the possibility of bad bandwagons. Similar analysis can illuminate the case in which states differ in the precision of their private information but not their size. In this case, we show that regardless of whether states allocate their delegates in direct proportion to the fraction of votes received by each candidate or they allocate all of their delegates to the candidate who received the majority of votes in that state, then ordering the states from most informed to least informed maximizes the probability that the higher quality candidate will be elected:

Proposition 6. *Suppose that $\pi_j^N=\pi^N$, $\lambda_j=\lambda$, $\kappa_j=0$, 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 states are ordered from most informed to least informed.*

4. Endogenous candidate strategies

In the previous sections 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 neutrals by the same amount.[11] 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 $\phi=\frac{1}{2}$ and that π_j^A 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 π_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 $\tilde{\pi}_j^A(\pi_j^A, a_j, b_j)$, where $\tilde{\pi}_j^A(\pi_j^A, a_j, b_j)$ is π_j^A if $a_j\leq b_j$, $\pi_j^A+a_j-b_j$ if $a_j\in(b_j, b_j+\pi_j^N]$, and $\pi_j^A+\pi_j^N$ if $a_j>b_j+\pi_j^N$. The fraction of B-partisans in state S_j is $\tilde{\pi}_j^B(\pi_j^B, a_j, b_j)$, where $\tilde{\pi}_j^B(\pi_j^B, a_j, b_j)$ is π_j^B if $b_j\leq a_j$, $\pi_j^B+b_j-a_j$ if $b_j\in(a_j, a_j+\pi_j^N]$, and $\pi_j^B+\pi_j^N$ if $b_j>a_j+\pi_j^N$. The fraction of neutral voters in state S_j is then just $\tilde{\pi}_j^N=1-\tilde{\pi}_j^A-\tilde{\pi}_j^B$. Voters then vote sincerely after candidates choose their campaign strategies.

Given these assumptions we obtain the following result:

Proposition 7. *Suppose there exists a pure strategy equilibrium to the candidate budget allocation game.[12] 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

[11] The assumption that campaign spending is used to swing neutral voters is standard in the literature. For example, Bombardini and Trebbi (2011) note that it is reasonable to model campaign spending as being used to sway neutrals to support a particular candidate.

[12] One can show that such an equilibrium exists with the benefit of additional assumptions.

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.[13] 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.

The proposition we have given is proven by showing that, under the conditions given in this section, 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 is unaffected by endogenous candidate strategies, the optimal order of the states 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 information about 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. We also show in a working paper version of this manuscript (Hummel and Holden, 2013) that these results extend to settings where one explicitly models the possibility that voters in some states care about different issues than the rest of the country. Thus our conclusions hold under a variety of scenarios that one might initially suspect could cause the optimal ordering of the states that we have proposed to go wrong.

[13] There 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.

Appendix A

Proof of Proposition 1. First note that if $(p(1-\kappa)+\kappa)\pi^N \geq \frac{1}{2}$, 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 $(p(1-\kappa)+\kappa)\pi^N < \frac{1}{2}$ 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 jth 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 $1-(p(1-\kappa)+\kappa)\pi^N$. 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 $1-(p(1-\kappa)+\kappa)\pi^N$. 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 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 $1-(p(1-\kappa)+\kappa)\pi^N$, 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 $1-(p(1-\kappa)+\kappa)\pi^N$. 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 greater if the order of the states $S_{(j)}$ and $S_{(j+1)}$ is reversed. First we show this for the case where the states allocate their delegates in direct proportion to the number of voters who voted for each candidate.

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 $1-(p(1-\kappa)+\kappa)\pi^N$ or the fraction of voters in state $S_{(j)}$ that votes for candidate A is less than or equal to $1-(p(1-\kappa)+\kappa)\pi^N$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is greater than $1-(p(1-\kappa)+\kappa)\pi^N$. Conditional on the fraction of voters in state $S_{(j)}$ that votes for candidate A being greater than $1-(p(1-\kappa)+\kappa)\pi^N$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ 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 $1-(p(1-\kappa)+\kappa)\pi^N$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A being greater than $1-(p(1-\kappa)+\kappa)\pi^N$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[(p(1-\kappa)+\kappa)\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate

A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$.

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

Now note that if $\pi^N \geq 1-(p(1-\kappa)+\kappa)\pi^N$, then the uniform distribution on $[\pi^N, 1]$ strictly first order stochastically dominates the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$, and the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ strictly first order stochastically dominates the uniform distribution on $[(p(1-\kappa)+\kappa)\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$. Thus if $\pi^N \geq 1-(p(1-\kappa)+\kappa)\pi^N$ 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 $\pi^N < 1-(p(1-\kappa)+\kappa)\pi^N$. Rewriting the distribution of the fraction of voters in state $S_{(j)}$ that votes for candidate A, we see that if the values of $\pi_{(k)}^A$ satisfy the conditions in Case 3, then with probability $\frac{1-(1+p(1-\kappa)+\kappa)\pi^N}{2-(1+2(p(1-\kappa)+\kappa))\pi^N}$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$. With probability $\frac{(p(1-\kappa)+\kappa)\pi^N}{2-(1+2(p(1-\kappa)+\kappa))\pi^N}$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1]$. With probability $\frac{(1-(p(1-\kappa)+\kappa))\pi^N}{2-(1+2(p(1-\kappa)+\kappa))\pi^N}$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[(p(1-\kappa)+\kappa)\pi^N, \pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$. And with probability $\frac{1-(1+p(1-\kappa)+\kappa)\pi^N}{2-(1+2(p(1-\kappa)+\kappa))\pi^N}$, the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$.

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 $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$ as it is that the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[\pi^N, 1-(p(1-\kappa)+\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$. 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 are if the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1]$ or if the fraction of voters in state $S_{(j)}$ that votes for candidate A is drawn from the uniform distribution on $[(p(1-\kappa)+\kappa)\pi^N, \pi^N]$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is drawn from the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$.

Now the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1]$ strictly first order stochastically dominates the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$. And the uniform distribution on $[1-(p(1-\kappa)+\kappa)\pi^N, 1-(1-p)(1-\kappa)\pi^N]$ strictly first order stochastically dominates the uniform distribution on $[(p(1-\kappa)+\kappa)\pi^N, \pi^N]$. 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 $\pi^N > 1-(p(1-\kappa)+\kappa)\pi^N$, 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 if the states allocate their delegates in direct proportion to the number of votes the candidates received.

Now we show that reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the probability that candidate A is elected if the states allocate all of their delegates to the candidate who received the most votes in each state. To see this, recall that if the values of $\pi_{(k)}^A$ satisfy the conditions in Case 3, then either the fraction of voters in state $S_{(j)}$ that votes for candidate A is greater than $1-(p(1-\kappa)+\kappa)\pi^N$ or the fraction of voters in state $S_{(j)}$ that votes for candidate A is less than or equal to $1-(p(1-\kappa)+\kappa)\pi^N$ and the fraction of voters in state $S_{(j+1)}$ that votes for candidate A is greater than $1-(p(1-\kappa)+\kappa)\pi^N$. In the first of these scenarios, a bandwagon begins in state $S_{(j)}$, states $S_{(j)}$ and $S_{(j+1)}$ both allocate all of their delegates to candidate A, and the order of these states does not affect the number of delegates that candidate A receives in these states. In the second of these scenarios, a bandwagon begins in state $S_{(j+1)}$, so the $j+1$th state allocates all of its delegates to candidate A, but the jth state might not allocate its delegates to candidate A. In this scenario, candidate A receives at least as many delegates in these states if the order of the states $S_{(j)}$ and $S_{(j+1)}$ is reversed.

But in either scenario, reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the probability that candidate A is elected. Thus regardless of how the states allocate their delegates, ordering the states from smallest to largest maximizes the probability that the higher quality candidate wins the election. □

Proof of Proposition 2. First note that if $\left(p_j(1-\kappa_j)+\kappa_j\right)\pi^N \geq \frac{1}{2}$ 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 $\left(p_j(1-\kappa_j)+\kappa_j\right)\pi^N \geq \frac{1}{2}$ 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 $\left(p_j(1-\kappa_j)+\kappa_j\right)\pi^N < \frac{1}{2}$ 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 $1 - (p_{(k)}(1-\kappa_{(k)}) + \kappa_{(k)})\pi^N$. 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 $1 - (p_{(k)}(1-\kappa_{(k)}) + \kappa_{(k)})\pi^N$. 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 number of delegates in states $S_{(j)}$ and $S_{(j+1)}$ for candidate *A* when the order of these states is reversed strictly first order stochastically dominates the distribution of the total number of delegates for candidate *A* in these states under the original order. First we prove this when states allocate their delegates in direct proportion to the number of voters who voted for the candidates in each state.

To see this, note that if $0 \leq \pi^A_{(j)} \leq 1 - 2(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N$, then neutral voters in the $j+1$ th state to vote according to their private signals regardless of the order in which states vote. Thus if $0 \leq \pi^A_{(j)} \leq 1 - 2(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N$, then the total number of voters who vote for *A* in states $S_{(j)}$ and $S_{(j+1)}$ is $\lambda\{\pi^A_{(j)} + (p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N + \pi^A_{(j+1)} + (p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N\}$ regardless of the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote, and the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote has no effect on the probability with which the candidates are elected.

If $1 - 2(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N < \pi^A_{(j)} \leq 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then neutral voters in the $j+1$th state vote vote according to their private signals if the order states $S_{(j)}$ and $S_{(j+1)}$ is not reversed, but vote for candidate *A* if the order of the states is reversed. Thus if $1 - 2(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N < \pi^A_{(j)} \leq 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then the total number of voters who vote for *A* in states $S_{(j)}$ and $S_{(j+1)}$ is $\lambda\{\pi^A_{(j)} + (p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N + \pi^A_{(j+1)} + (p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N\}$ if the order of the states is not reversed and $\lambda\{\pi^A_{(j)} + (p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N + \pi^A_{(j+1)} + \pi^N\}$ if the order of the states is reversed. Thus if $1 - 2(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N < \pi^A_{(j)} \leq 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then candidate *A* receives a strictly larger number of votes if the order of the states is reversed.

And if $\pi^A_{(j)} > 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then neutral voters in the $j+1$th state to vote vote for candidate *A* regardless of the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote. Thus if $\pi^A_{(j)} > 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then the total number of voters who vote for candidate *A* in states $S_{(j)}$ and $S_{(j+1)}$ is $\lambda\{\pi^A_{(j)} + (p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N + \pi^A_{(j+1)} + \pi^N\}$ if the order of the states is not reversed and $\lambda\{\pi^A_{(j)} + (p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)})\pi^N + \pi^A_{(j+1)} + \pi^N\}$ if the order of the states is reversed. Thus if $\pi^A_{(j)} > 1 - 2(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)})\pi^N$, then candidate *A* receives a strictly larger number of votes if the order of the states is reversed.

In any of these cases, the total number of voters who vote for *A* is at least as large and sometimes strictly larger if the order of states $S_{(j)}$ and $S_{(j+1)}$ is reversed. 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 if delegates are allocated in direct proportion to the number of voters who voted for the candidates in each state.

Now we show that reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the probability that candidate *A* is elected if each state allocates all of its delegates to the candidate who received the most votes in that state. To see this, first note that the probability a bandwagon begins for candidate *A* if voters are voting according to their private signals is greater in state $S_{(j+1)}$ than in state $S_{(j)}$. Also note that the probability that state $S_{(j)}$ would not allocate its delegates to candidate *A* is greater than the probability that state $S_{(j+1)}$ would not allocate its delegates to candidate *A* if neutral voters vote according to their private signals.

Combining the facts in the previous paragraph shows that reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote both increases the probability that there will be a bandwagon for candidate *A* in the *j*th state to vote and also increases the probability that this will change whether the $j+1$th state to vote allocates its delegates to candidate *A*. Thus reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the probability that both of these states will allocate all of their delegates to candidate *A*.

At the same time, note that states $S_{(j)}$ and $S_{(j+1)}$ will allocate all of their delegates to candidate *B* if and only if $y_{(j)} \leq \frac{1}{2}$ and $y_{(j+1)} \leq \frac{1}{2}$. This happens with probability $F\left(\frac{1}{2} - \left(p_{(j)}(1-\kappa_{(j)}) + \kappa_{(j)}\right)\pi^N\right) F\left(\frac{1}{2} - \left(p_{(j+1)}(1-\kappa_{(j+1)}) + \kappa_{(j+1)}\right)\pi^N\right)$ regardless of the order of states $S_{(j)}$ and $S_{(j+1)}$, where *F* denotes the cumulative distribution function satisfying $F = F_j$ for all *j*. Thus having the state with the more accurate private signals vote first does not affect the probability that both of these states will allocate all of their delegates to candidate *B* and it makes it more likely that both of these states will allocate all of their delegates to candidate *A*. This implies that reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the probability that candidate *A* is elected if states allocate all of their delegates to the candidate who received the most votes in that state. □

Proof of Proposition 3. Since $\delta < \max\{((2p-1)(1-\kappa) + \kappa)\pi^N, \frac{1}{2} - (p + (1-p)\kappa)\pi^N\}$, it follows that if $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N \in [0, \delta]$, then $y_{(1)} - (1-p)(1-\kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$ and if $y_{(1)} - (1-p)(1-\kappa)\pi^N \in [1 - \pi^N - \delta, 1 - \pi^N]$, then $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$. From the assumptions on $f(x)$ we know that $\phi_{(2)} < 1 - p$ if and only if $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N < \delta$ and $y_{(1)} - (1-p)(1-\kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$ or $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N < 0$. Combining this with the first result in this proof indicates that $\phi_{(2)} < 1 - p$ if and only if $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N < \delta$ or $y_{(1)} < \delta + (p(1-\kappa) + \kappa)\pi^N$.

Similarly, from the assumptions on $f(x)$ we know that $\phi_{(2)} > p$ if and only if either $y_{(1)} - (1-p)(1-\kappa)\pi^N > 1 - \pi^N - \delta$ and $y_{(1)} - (p(1-\kappa) + \kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$ or $y_{(1)} - (1-p)(1-\kappa)\pi^N > 1 - \pi^N$. Combining this with the first result in this proof indicates that $\phi_{(2)} > p$ if and only if $y_{(1)} - (1-p)(1-\kappa)\pi^N > 1 - \pi^N - \delta$ or $y_{(1)} > 1 - \delta - (p(1-\kappa) + \kappa)\pi^N$.

Now suppose the state of the world is *a*. In this case $\pi^A_{(1)} = y_{(1)} - (p(1-\kappa) + \kappa)\pi^N$. Combining this with the results in the previous two paragraphs shows that $\phi_{(2)} < 1 - p$ if and only if $\pi^A_{(1)} < \delta$ and $\phi_{(2)} > p$ if and only if $\pi^A_{(1)} > 1 - \delta - 2(p(1-\kappa) + \kappa)\pi^N$. Thus imperfectly informed neutral voters in the second state vote for B if and only if $\pi^A_{(1)} < \delta$, imperfectly informed neutral voters in the second state vote for A if and only if $\pi^A_{(1)} > 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, and all neutral voters vote according to their private signals if and only if $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$.

Note that if $\pi^A_{(1)} < \delta$ and all imperfectly informed 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 $\pi^A_{(1)} > 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$ and all imperfectly informed 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 $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$. We first show that when the states allocate their votes in direct proportion to the number of votes each candidate received, then candidate A is elected with greater probability when $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$ if the small state votes first.

If $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, then the total number of voters who vote for A is $\lambda_{(1)}(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N) + \lambda_{(2)}(\pi^A_{(2)} + (p(1-\kappa)+\kappa)\pi^N)$. Thus candidate A wins the election if and only if $\lambda_{(1)}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N\right) + \lambda_{(2)}\left(\pi^A_{(2)} + (p(1-\kappa)+\kappa)\pi^N\right) \ge \frac{1}{2}(\lambda_{(1)} + \lambda_{(2)})$ or $\lambda_{(1)}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right) + \lambda_{(2)}\left(\pi^A_{(2)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right) \ge 0$ or $\frac{\lambda_{(1)}}{\lambda_{(2)}}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right) + \pi^A_{(2)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2} \ge 0$ or $\pi^A_{(2)} \ge \frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_{(1)}}{\lambda_{(2)}}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right)$. So for a fixed $\pi^A_{(1)}$ satisfying $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, candidate A wins the election with probability $1 - F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_{(1)}}{\lambda_{(2)}}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right)\right)$, where I let F denote the cumulative distribution function satisfying $F = F_1 = F_2$.

From this it follows that if $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, then candidate A wins the election with probability $\int_{\delta}^{1-\delta-2(p(1-\kappa)+\kappa)\pi^N}\left[1 - F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_{(1)}}{\lambda_{(2)}}\left(\pi^A_{(1)} + (p(1-\kappa)+\kappa)\pi^N - \frac{1}{2}\right)\right)\right] f\left(\pi^A_{(1)}\right) d\pi^A_{(1)} / Pr\left(\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N\right) = \int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N}\left[1 - F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_{(1)}}{\lambda_{(2)}}x\right)\right] f\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + x\right)dx / Pr\left(\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N\right)$. 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 if and only if $\int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right) f\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + x\right)dx \le \int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right) f\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + x\right)dx$. And since $f(x)$ is constant for all $x \in [\delta, 1 - \pi^N - \delta]$, it follows that candidate A wins with greater probability when the small state votes first if and only if $\int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right)dx \le \int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right)dx$.

Now $\int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right)dx = \int_0^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_1}{\lambda_2}x\right)dx$ and $\int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right)dx = \int_0^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_2}{\lambda_1}x\right)dx$. And $\frac{d}{du}\left[F\left(\frac{1}{2} - p\pi^N - u\right) + F\left(\frac{1}{2} - p\pi^N + u\right)\right] = \left[f\left(\frac{1}{2} - p\pi^N + u\right) - f\left(\frac{1}{2} - p\pi^N - u\right)\right] \ge 0$ since $f(x)$ is symmetric and weakly single-peaked about $x = \frac{1-\pi^N}{2} \ge \frac{1}{2} - p\pi^N$. Thus since $\lambda_2 > \lambda_1$, $F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_2}{\lambda_1}x\right)$ is at least as large as $F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_1}{\lambda_2}x\right)$ for any given $x > 0$. From this it follows that $\int_0^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_1}{\lambda_2}x\right)dx \le \int_0^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right) + F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N + \frac{\lambda_2}{\lambda_1}x\right)dx$ and $\int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_1}{\lambda_2}x\right)dx \le \int_{\delta+(p(1-\kappa)+\kappa)\pi^N-\frac{1}{2}}^{\frac{1}{2}-\delta-(p(1-\kappa)+\kappa)\pi^N} F\left(\frac{1}{2} - (p(1-\kappa)+\kappa)\pi^N - \frac{\lambda_2}{\lambda_1}x\right)dx$.

Thus when $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, candidate A wins with greater probability when the small state votes first if the states allocate their delegates in direct proportion to the number of votes each candidate received. We now show that when the states allocate all of their delegates to the candidate who received the most votes in the state, then candidate A is elected with greater probability when $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$ if the small state votes first.

To see this, note that if each state allocates all of its delegates to the candidate who receives the most votes in the state, then candidate A wins the election if and only if the majority of voters in the large state vote for candidate A. Now if the large state votes first, then the probability that the majority of voters in the large state votes for candidate A, conditional on $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, is equal to the probability that the majority of voters in the large state votes for candidate A, conditional on $\delta + (p(1-\kappa)+\kappa)\pi^N \le y^A_{(1)} \le 1 - \delta - (p(1-\kappa)+\kappa)\pi^N$, or $\frac{1}{2}$. But if the large state votes second, then the probability that the majority of voters in the large state votes for candidate A, conditional on $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p(1-\kappa)+\kappa)\pi^N$, is just the unconditional probability that the majority of voters in the large state votes for candidate A, which is greater than $\frac{1}{2}$.

Thus candidate A wins with greater probability under this scenario when the small state votes first if the states allocate all of their delegates to the candidate who received the most votes in the state, and the probability that A is elected is always maximized by having the small state vote first. 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. □

Proof of Proposition 4. Since $\delta < \max\left\{\left(\left(2p_j - 1\right)(1-\kappa) + \kappa\right)\pi^N, \frac{1}{2} - \left(p_j + \left(1 - p_j\right)\kappa\right)\pi^N\right\}$ for $j = 1$ and 2, it follows that if $y_j - (p_j(1-\kappa)+\kappa)\pi^N \in [0, \delta]$, then $y_j - (1-p_j)(1-\kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$ and if $y_j - (1-p_j)(1-\kappa)\pi^N \in [1 - \pi^N - \delta, 1 - \pi^N]$, then $y_j - (p_j(1-\kappa)+\kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$. From the assumptions on $f(x)$ we know that $\phi_{(2)} < 1 - p_{(2)}$ if and only if $y_{(1)} - (p_{(1)}(1-\kappa)+\kappa)\pi^N < \delta$ and $y_{(1)} - (1-p_{(1)})(1-\kappa)\pi^N \in [\delta, 1 - \pi^N - \delta]$ or $y_{(1)} - (p_{(1)}(1-\kappa)+\kappa)\pi^N < 0$. Combining this with the first result in this proof indicates that $\phi_{(2)} > p_{(2)}$ if and only if $y_{(1)} - (p_{(1)}(1-\kappa)+\kappa)\pi^N < \delta$ or $y_{(1)} < \delta + (p_{(1)}(1-\kappa)+\kappa)\pi^N$.

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

Now suppose that the state of the world is a. In this case $\pi^A_{(1)} = y_{(1)} - (p_{(1)}(1-\kappa)+\kappa)\pi^N$. Combining this with the results in the previous two paragraphs shows that $\phi_{(2)} < 1 - p_{(2)}$ if and only if $\pi^A_{(1)} < \delta$ and $\phi_{(2)} > p_{(2)}$ if and only if $\pi^A_{(1)} > 1 - \delta - 2(p_{(1)}(1-\kappa)+\kappa)\pi^N$. Thus all imperfectly informed neutral voters in the second state vote for B if and only if $\pi^A_{(1)} < \delta$, all imperfectly informed neutral voters in the second state vote for A if and only if $\pi^A_{(1)} > 1 - \delta - 2(p_{(1)}(1-\kappa)+\kappa)\pi^N$, and all neutral voters vote according to their private signals if and only if $\delta \le \pi^A_{(1)} \le 1 - \delta - 2(p_{(1)}(1-\kappa)+\kappa)\pi^N$.

From this it follows that if $\pi^A_{(1)} < \delta$, then the total number of voters who vote for A is $\lambda(\pi^A_{(1)} + (p_{(1)}(1-\kappa)+\kappa)\pi^N + \pi^A_{(2)} + \kappa\pi^N)$. If

$\delta \leq \pi^A_{(1)} \leq 1 - \delta - 2\max\{p_{(1)}(1-\kappa)+\kappa, p_{(2)}(1-\kappa)+\kappa\}\pi^N$, then the total number of voters who vote for A is $\lambda(\pi^A_{(1)} + (p_{(1)}(1-\kappa)+\kappa)\pi^N + \pi^A_{(2)} + (p_{(2)}(1-\kappa)+\kappa)\pi^N)$. If $1-\delta-2\max\{p_{(1)}(1-\kappa)+\kappa, p_{(2)}(1-\kappa)+\kappa\}\pi^N < \pi^A_{(1)} \leq 1-\delta-2\min\{p_{(1)}(1-\kappa)+\kappa, p_{(2)}(1-\kappa)+\kappa\}\pi^N$, then the total number of voters who vote for A is $\lambda(\pi^A_{(1)} + (p_{(1)}(1-\kappa)+\kappa)\pi^N + \pi^A_{(2)} + (p_{(2)}(1-\kappa)+\kappa)\pi^N)$ if $p_{(1)} < p_{(2)}$ and $\lambda(\pi^A_{(1)} + (p_{(1)}(1-\kappa)+\kappa)\pi^N + \pi^A_{(2)} + \pi^N)$ otherwise. And if $\pi^A_{(1)} > 1-\delta-2\min\{p_{(1)}(1-\kappa)+\kappa, p_{(2)}(1-\kappa)+\kappa\}\pi^N$, then the total number of voters who vote for A is $\lambda(\pi^A_{(1)} + (p_{(1)}(1-\kappa)+\kappa)\pi^N + \pi^A_{(2)} + \pi^N)$.

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 signals votes first. Thus when states allocate their delegates in direct proportion to the number of voters who voted for the candidates, the probability that A is elected is maximized when the state with the more accurate private signals votes first.

Now we show that if each state allocates all of its delegates to the candidate who received the most votes in that state, then the probability that A is elected is maximized when the state with the more accurate private signals votes first. To see this, first note that the probability that a bandwagon begins for candidate A if voters are voting according to their private signals is greater if the state with the more accurate private signals votes first. Also note that the probability that a state would not allocate its delegates to candidate A if neutral voters vote according to their private signals is greater when the voters in the state have less accurate private information.

Combining the facts in the previous paragraph shows that when the state with the more accurate private information votes first, there is a greater probability that there will be a bandwagon for candidate A in the first state, and there is also a greater probability that this bandwagon will cause the second state to change from allocating all of its delegates to candidate B to allocating all of its delegates to candidate A. Thus having the state with the more accurate private signals vote first makes it more likely that both states will allocate all of their delegates to candidate A.

At the same time, note that both states will allocate all of their delegates to candidate B if and only if either $\pi^A_{(1)} < \delta$ or $\delta \leq \pi^A_{(1)} \leq \frac{1}{2} - \left(p_{(1)}(1-\kappa)+\kappa\right)\pi^N$ and $\pi^A_{(2)} \leq \frac{1}{2} - \left(p_{(2)}(1-\kappa)+\kappa\right)\pi^N$. This happens with probability $F(\delta) + \left[F\left(\frac{1}{2} - \left(p_{(1)}(1-\kappa)+\kappa\right)\pi^N\right) - F(\delta)\right] F\left(\frac{1}{2} - \left(p_{(2)}(1-\kappa)+\kappa\right)\pi^N\right) = F\left(\frac{1}{2} - \left(p_{(1)}(1-\kappa)+\kappa\right)\pi^N\right) F\left(\frac{1}{2} - \left(p_{(2)}(1-\kappa)+\kappa\right)\pi^N\right) + F(\delta)\left[1 - F\left(\frac{1}{2} - \left(p_{(2)}(1-\kappa)+\kappa\right)\pi^N\right)\right]$, where F denotes the cumulative distribution function satisfying $F = F_1 = F_2$. This probability is smaller when the state with the more accurate private signals votes first. Thus having the state with the more accurate private signals vote first makes it less likely that both states will allocate all of their delegates to candidate B, and having the state with the more accurate private signals vote first increases the probability that candidate A is elected.

These results show that when the state of the world is a, then either method for allocating delegates increases the probability that candidate A is elected. 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 signals votes first. Thus the probability that the better candidate is elected is maximized when the state with the more accurate private signals votes first. □

Proof of Proposition 5. By the same reasoning presented at the start of the proof of Proposition 3 applied to the special case in which $\kappa = 0$, it follows that if neutral voters in state $S_{(j)}$ vote according to their private signals, then we have $\phi_{(j+1)} = \phi_{(j)}$ if $y_{(j)} \in [\delta + p\pi^N, 1 - \delta - p\pi^N]$, $\phi_{(j+1)} > p$ if $y_{(j)} > 1 - \delta - p\pi^N$, and $\phi_{(j+1)} < 1 - p$ if $y_{(j)} < \delta + p\pi^N$. From this it follows that if neutral voters in state $S_{(j)}$ vote according to their private signals, then neutral voters in state $S_{(j+1)}$ vote according to their private signals if $y_{(j)} \in [\delta + p\pi^N, 1 - \delta - p\pi^N]$, neutral voters in state $S_{(j+1)}$ vote for candidate B if $y_{(j)} < \delta + p\pi^N$, and neutral voters in state $S_{(j+1)}$ vote for candidate A if $y_{(j)} > 1 - \delta - p\pi^N$. Furthermore, if neutral voters in a given state all vote for the same candidate regardless of their private signals, then voters in future states do not update their beliefs about the identity of the higher quality candidate from the results of this earlier state.

By combining the insights in the previous paragraph, it follows that under the assumptions in this proposition, neutral voters in each 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 $1 - \delta - p\pi^N$ in some previous state $S_{(k)}$ with $k < j$, then neutral voters in state $S_{(j)}$ vote for that particular candidate. And if no candidate has received a fraction of the vote greater than $1 - \delta - p\pi^N$ in some previous state $S_{(k)}$ with $k < j$, then neutral voters in state $S_{(j)}$ vote according to their private signals.

Now 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 $S_{(1)}, \ldots, S_{(m)}$ that is 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 the 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 jth state to vote, we consider several 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 a particular candidate is greater than $1 - \delta - p\pi^N$. 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 a particular candidate is greater than $1 - \delta - p\pi^N$. 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. 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 a particular candidate is greater than $1 - \delta - p\pi^N$, but the value of $\pi^A_{(j)}$ is such that the fraction of voters in state $S_{(j)}$ that votes for a particular candidate is greater than $1 - \delta - p\pi^N$. In this case, the majority of voters in states $S_{(k)}$ with $k \geq j$ vote for the candidate that received the majority of votes in state $S_{(j)}$ regardless of the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote, and reversing the order of these two states again has no effect on the probability with which candidate A is elected.

Case 4. If none of the above three possibilities holds, then the values of $\pi^A_{(k)}$ for $k = 1, \ldots, j+1$ are such that there is no state $S_{(k)}$ with $k \leq j$ for which the fraction of voters in state $S_{(k)}$ that votes for a particular candidate is greater than $1 - \delta - p\pi^N$, but the value of $\pi^A_{(j+1)}$ is such that the fraction of voters in state $S_{(j+1)}$ that votes for a particular candidate is greater than $1 - \delta - p\pi^N$. Now conditional on the fraction of voters in state $S_{(k)}$ that votes for a particular candidate being no greater than $1 - \delta - p\pi^N$, then the fraction of voters that votes for candidate A is a random variable drawn from the uniform distribution on $[\delta + p\pi^N, 1 - \delta - p\pi^N]$. From this it follows that conditional on the fraction of voters in state $S_{(k)}$ that votes for a particular candidate being no greater than $1 - \delta - p\pi^N$, the probability that the majority of voters in state $S_{(k)}$ votes for candidate A is $\frac{1}{2}$.

Similarly, given that the fraction of voters in state $S_{(j+1)}$ that votes for a particular candidate is greater than $1-\delta-p\pi^N$, all neutral voters in states $S_{(k)}$ with $k>j+1$ will vote for the candidate that received the majority of votes in state $S_{(j+1)}$. Thus conditional on the values of $\pi^A_{(k)}$ for $k=1,\ldots,j+1$ satisfying the conditions in this case, the probability that the majority of voters in state $S_{(k)}$ vote for candidate A conditional on the majority of voters in state $S_{(j+1)}$ voting for candidate A is the same as the probability that the majority of voters in state $S_{(k)}$ vote for candidate B conditional on the majority of voters in state $S_{(j+1)}$ voting for candidate B for all $k\neq j+1$. From this it follows that the distribution of delegates that are cast for candidate A conditional on the majority of voters in state $S_{(j+1)}$ voting for candidate A is the same as the distribution of delegates that are cast for candidate B conditional on the majority of voters in state $S_{(j+1)}$ voting for candidate B.

Now changing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote by making the larger state vote later can only increase the probability that the candidate who receives the majority of votes in state $S_{(j+1)}$ will be elected. And conditional on the fraction of voters in state $S_{(j+1)}$ that votes for candidate A being greater than $1-\delta-p\pi^N$, it is equally likely that changing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote will change the outcome of the election to candidate A being elected as it is that, conditional on the fraction of voters in state $S_{(j+1)}$ that votes for candidate B being greater than $1-\delta-p\pi^N$, changing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote will change the outcome of the election to candidate B being elected. But it is more likely that the fraction of voters in state $S_{(j+1)}$ that votes for the higher quality candidate will be greater than $1-\delta-p\pi^N$ than it is that the fraction of voters in state $S_{(j+1)}$ that votes for the lower quality candidate will be greater than $1-\delta-p\pi^N$. By combining these facts it then follows that changing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote by making the larger state vote later increases the probability that the higher quality candidate is elected. The result then follows. □

Proof of Proposition 6. By the same reasoning presented at the start of the proof of Proposition 4 applied to the special case in which $\kappa=0$, it follows that if neutral voters in state $S_{(j)}$ vote according to their private signals, then we have $\phi_{(j+1)}=\phi_{(j)}$ if $y_{(j)}\in[\delta+p_{(j)}\pi^N,1-\delta-p_{(j)}\pi^N]$, $\phi_{(j+1)}>p_{(j+1)}$ if $y_{(j)}>1-\delta-p_{(j)}\pi^N$, and $\phi_{(j+1)}<1-p_{(j+1)}$ if $y_{(j)}<\delta+p_{(j)}\pi^N$. From this it follows that if neutral voters in state $S_{(j)}$ vote according to their private signals, then neutral voters in state $S_{(j+1)}$ vote according to their private signals if $y_{(j)}\in[\delta+p_{(j)}\pi^N,1-\delta-p_{(j)}\pi^N]$, neutral voters in state $S_{(j+1)}$ vote for candidate B if $y_{(j)}<\delta+p_{(j)}\pi^N$, and neutral voters in state $S_{(j+1)}$ vote for candidate A if $y_{(j)}>1-\delta-p_{(j)}\pi^N$. Furthermore, if neutral voters in a given state all vote for the same candidate regardless of their private signals, then voters in future states do not update their beliefs about the identity of the higher quality candidate from the results of this earlier state.

By combining the insights in the previous paragraph, it follows that under the assumptions in this proposition, neutral voters in each 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 $1-\delta-p_{(k)}\pi^N$ in some previous state $S_{(k)}$ with $k<j$, then neutral voters in state $S_{(j)}$ vote for that particular candidate. And if no candidate has received a fraction of the vote greater than $1-\delta-p_{(k)}\pi^N$ in some previous state $S_{(k)}$ with $k<j$, then neutral voters in state $S_{(j)}$ vote according to their private signals.

Now 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 $S_{(1)},\ldots,S_{(m)}$ that is 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 the 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 jth state to vote, we consider several 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 a particular candidate is greater than $1-\delta-p_{(k)}\pi^N$. 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 a particular candidate is greater than $1-\delta-p_{(k)}\pi^N$. We show that, conditional on the values of $\pi^A_{(k)}$ for $k=1,\ldots,j-1$ satisfying these conditions, the probability that candidate A will be elected is strictly greater when the order of these states is reversed by illustrating that the distribution of the total number of delegates for candidate A first order stochastically dominates the distribution of the total number of delegates for candidate A under the original order. First we prove this when states allocate their delegates in direct proportion to the number of voters who voted for the candidates in each state.

To see this, note that if $\pi^A_{(j)}\leq\delta$, then neutral voters in all states after the jth state all vote for candidate B regardless of the order in which the states vote, so reversing the order in which the states vote changes the total number of voters who vote for A in states $S_{(j)}$ and $S_{(j+1)}$ from $\lambda(\pi^A_{(j)}+p_{(j)}\pi^N+\pi^A_{(j+1)})$ to $\lambda(\pi^A_{(j)}+p_{(j+1)}\pi^N+\pi^A_{(j+1)})$ while having no effect on how voters in states $S_{(k)}$ with $k<j$ vote. Thus in this case, reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the fraction of voters who vote for candidate A.

Also note that if $1-\delta-2p_{(j+1)}\pi^N\leq\pi^A_{(j)}$, then neutral voters in all states after the jth state all vote for candidate A if the order of the states is reversed and at the same time reversing the order of the states increases the number of voters who vote for candidate A in the jth state to vote from $\lambda(\pi^A_{(j)}+p_{(j)}\pi^N)$ to $\lambda(\pi^A_{(j)}+p_{(j+1)}\pi^N)$. Thus reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote increases the fraction of voters who vote for candidate A in this case.

Next note that if $\delta\leq\pi^A_{(j)}\leq1-\delta-2p_{(j+1)}\pi^N$ and either $\pi^A_{(j+1)}\leq1-\delta-2p_{(j+1)}\pi^N$ or $\pi^A_{(j+1)}\geq1-\delta-2p_{(j)}\pi^N$, then neutral voters in states $S_{(j)}$ and $S_{(j+1)}$ vote according to their private signals regardless of the order of these states, and reversing the order in which these states vote also has no effect on how neutral voters in states $S_{(k)}$ with $k>j+1$ will vote. Thus reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote has no effect on the number of votes received by each candidate in this case.

Finally consider what happens if $\delta\leq\pi^A_{(j)}\leq1-\delta-2p_{(j+1)}\pi^N$ and $1-\delta-2p_{(j+1)}\pi^N\leq\pi^A_{(j+1)}\leq1-\delta-2p_{(j)}\pi^N$. In this circumstance, reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote would cause neutral voters in states $S_{(k)}$ with $k>j+1$ to vote the same way they would vote in the absence of any information from these states rather than all voting for candidate A while having no effect on the total number of voters who vote for A in states $S_{(j)}$ and $S_{(j+1)}$ (which is equal to $\lambda(\pi^A_{(j)}+p_{(j)}\pi^N+\pi^A_{(j+1)}+p_{(j+1)}\pi^N)$ regardless of the order of the states). However, if $\pi^A_{(j)}$ were instead equal to the value of $\pi^A_{(j+1)}$ from this example and $\pi^A_{(j+1)}$ were equal to the value of $\pi^A_{(j)}$ from this example, then reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote would have the opposite effect on how neutral voters in states $S_{(k)}$ with $k>j+1$ vote while simultaneously increasing the total number of voters who vote for A in states $S_{(j)}$ and $S_{(j+1)}$ from $\lambda(\pi^A_{(j)}+p_{(j)}\pi^N+\pi^A_{(j+1)}+p_{(j+1)}\pi^N)$ to $\lambda(\pi^A_{(j)}+p_{(j+1)}\pi^N+\pi^A_{(j+1)}+\pi^N)$. Thus conditional on either $\delta\leq\pi^A_{(j)}\leq1-\delta-2p_{(j+1)}\pi^N$ and $1-\delta-$

$2p_{(j+1)}\pi^N \leq \pi^A_{(j+1)} \leq 1-\delta-2p_{(j)}\pi^N$ or $1-\delta-2p_{(j+1)}\pi^N \leq \pi^A_{(j+1)} \leq 1-\delta-2p_{(j+1)}\pi^N$ and $\delta \leq \pi^A_{(j+1)} \leq 1-\delta-2p_{(j+1)}\pi^N$ being satisfied, reversing the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote means that the resulting distribution of the number of voters who vote for candidate *A* first order stochastically dominates the original distribution of the number of voters who vote for candidate *A*.

Putting this all together shows that if the order in which states $S_{(j)}$ and $S_{(j+1)}$ vote is reversed, then the distribution of the number of voters who vote for candidate *A* strictly first order stochastically dominates the distribution of the number of voters who voted for candidate *A* under the original order. Thus the distribution of the total number of delegates for candidate *A* first order stochastically dominates the distribution of the total number of delegates for candidate *A* under the original order when states allocate their delegates in direct proportion to the number of voters who voted for the candidates in each state.

A substantively identical argument shows that the distribution of the total number of delegates for candidate *A* first order stochastically dominates the distribution of the total number of delegates for candidate *A* under the original order when states allocate all of their delegates to whichever candidate received the majority of votes in that state. From this it follows that ordering the states from most informed to least informed maximizes the probability that candidate *A* will be elected. □

Proof of Proposition 7. 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 σ_A and candidate *B* uses the strategy σ_B, then $\sigma_A = a^*$ maximizes the value of $\min_{\sigma_B} Pr(A \text{ wins}|(\sigma_A, \sigma_B))$.

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 σ_A and candidate *B* uses the strategy σ_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 π_j as it is that the fraction of *B*-partisans in state S_j before the candidates campaign is π_j. Thus it is equally likely that the fractions of *A* and *B*-partisans in state S_j after the candidates campaign are $\tilde{\pi}^A_j$ and $\tilde{\pi}^B_j$ respectively 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 fractions of *B* and *A*-partisans in state S_j after the candidates campaign are $\tilde{\pi}^A_j$ and $\tilde{\pi}^B_j$ respectively 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 fractions of *A* and *B*-partisans in state S_j after the candidates campaign are $\tilde{\pi}^A_j$ and $\tilde{\pi}^B_j$ respectively as it is that candidate *B* will win the election if the fractions of *B* and *A*-partisans in state S_j after the candidates campaign are $\tilde{\pi}^A_j$ and $\tilde{\pi}^B_j$ respectively. 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 $\min_{\sigma_A} Pr(B \text{ wins}|(\sigma_A, \sigma_B))$, 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

Abramson, P.R., Aldrich, J.H., Paolino, P., Rohde, D.W., 1992. "Sophisticated" voting in the 1988 presidential primaries. Am. Polit. Sci. Rev. 86, 55–69.

Aldrich, J.H., 1980. A dynamic model of presidential nomination campaigns. Am. Polit. Sci. Rev. 74, 651–669.

Ali, S.N., Kartik, N., 2012. Herding with collective preferences. Econ. Theory 51, 601–626.

Banerjee, A., 1992. A simple model of herd behavior. Q. J. Econ. 107, 797–818.

Bartels, L.M., 1985. Expectations and preferences in presidential nominating campaigns. Am. Polit. Sci. Rev. 79, 804–815.

Bartels, L.M., 1988. Presidential Primaries and the Dynamics of Public Choice. Princeton University Press, Princeton.

Battaglini, M., 2005. Sequential voting with abstention. Games Econ. Behav. 51, 445–463.

Battaglini, M., Morton, R., Palfrey, T., 2007. Efficiency, equity, and timing of voting mechanisms. Am. Polit. Sci. Rev. 101, 409–424.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom and cultural change as information cascades. J. Polit. Econ. 100, 992–1026.

Bombardini, M., Trebbi, F., 2011. Votes or money? Theory and evidence from the US congress. J. Public Econ. 95, 587–611.

Brams, S.J., Davis, M.D., 1982. Optimal resource allocation in presidential primaries. Math. Soc. Sci. 3, 373–388.

Buhr, T., 2000. What voters know about the candidates and how they learned it: the 1996 New Hampshire Republican primary as a case study. In: Mayer, W.G. (Ed.), In Pursuit of the White House: How We Choose Our Presidential Nominees. Chatham House Publishers, Chatham, pp. 203–254.

Callander, S., 2007. Bandwagons and momentum in sequential voting. Rev. Econ. Stud. 74, 653–684.

Chen, D., 2005. New Jersey moves to hold early primary. New York Times (June 24).

Dekel, E., Piccione, M., 2000. Sequential voting procedures in symmetric binary elections. J. Polit. Econ. 108, 34–55.

Deltas, G., Herrera, H., Polborn, M., 2012. A Theory of Learning and Coordination in the Presidential Primary System. University of Illinois Typescript.

Frazzini, A., 2006. The disposition effect and underreaction to news. J. Finance 61, 2017–2046.

Hirshleifer, S., Lim, S., Teoh, S.H., 2009. Driven to distraction: extraneous events and underreaction to earnings news. J. Finance 64, 2289–2325.

Hummel, P., Holden, R., 2013. Optimal primaries. NBER Working Paper No. 19340.

Hummel, P., Knight, B., 2012. Sequential or Simultaneous Elections? A Welfare Analysis. Brown University Typescript.

Ikenberry, D.L., Ramnath, S., 2002. Underrecation to self-selected news events. J. Finance 15, 489–526.

Kenney, P.J., Rice, T.W., 1994. The psychology of political momentum. Polit. Res. Q. 47, 923–938.

Klumpp, T., Polborn, M.K., 2006. Primaries and the New Hampshire effect. J. Public Econ. 90, 1073–1114.

Knight, B., Schiff, N., 2010. Momentum in presidential primaries. J. Polit. Econ. 118, 1110–1150.

Matthews, D.R., 1978. 'Winnowing': the news media and the 1976 presidential nominations. In: Barber, J.D. (Ed.), Race for the Presidency. Prentice-Hall Inc., Englewood Cliffs, pp. 55–78.

Mayer, W.G., Busch, A.E., 2004. The Front-loading Problem in Presidential Nominations. Brookings Institution Press, Washington, DC.

Norrader, B., 1992. Super Tuesday: Regional Politics and Presidential Primaries. University Press of Kentucky, Lexington.

Popkin, S.L., 1994. The Reasoning Voter: Communication and Persuasion in Presidential Campaigns. University of Chicago Press, Chicago.

Redslawk, D., Tolbert, C., Donavan, T., 2011. Why Iowa? How Caucuses and Sequential Elections Improve the Presidential Nomination Process. University of Chicago Press, Chicago.

Selman, D., 2010. Optimal Sequencing of Presidential Primaries. Bo aziçi University Typescript.

Strumpf, K.S., 2002. Strategic competition in sequential election contests. Public Choice 111, 377–397.

Vavreck, L., Spiliotes, C.J., Fowler, L.L., 2002. The effects of retail politics in the New Hampshire primary. Am. J. Polit. Sci. 46, 595–610.

von Neumann, J., 1928. Zur Theorie der Gesellschaftspielle. Math. Ann. 100, 295–320.