Primaries as Coordination Devices

Mattias Polborn*   Helios Herrera†

First draft: December 2007
This version: September 27, 2008

Abstract

To analyze the advantages and disadvantages of the U.S. presidential primary system, we construct a model in which three candidates compete for the nomination and voters are uncertain about the candidates' valence.

The disadvantage of simultaneous elections is that, if two of the candidates offer the same policy, then they split the voters with a policy preference for their position, which may allow the third candidate to win even if he is not the Condorcet winner. In contrast, in a sequential system, voters in late districts can use the election outcome of early primaries as a coordination device. Against this advantage stands the cost that coordination may occur on the wrong candidate. We find conditions under which a sequential system dominates a simultaneous system, and vice versa.

JEL Classification Numbers: D72, D60.
Keywords: Voting, primary elections, simultaneous versus sequential elections.

*Corresponding author. Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, IL 61801 USA, E-mail: polborn@uiuc.edu.

†SIPA, Columbia University and CIE-ITAM, Mexico, D.F. Email: hhh2108@columbia.edu

1 Introduction

Candidates for the U.S. presidential election are determined through a sequence of elections within each political party, the primaries.[1] The nomination process is one of the most controversial institutions of America's contemporary political landscape. The most common ground for attack on the modern primary system is the perception that its sequential structure is inherently "unfair" in that it shifts too much power to voters in early primary states.

Over the last several election cycles, many states have shifted their primaries earlier and earlier, while the national parties have tried to steer against this movement. For the 2008 cycle, both the Democratic and the Republican National Committee chose rules that prohibited all but a few states to hold their primaries before February 5th. Florida and Michigan violated these rules and were punished by the DNC and RNC by taking away half of their delegates at the convention, respectively.[2] This interesting fact illustrates two points: First, states have a strong interest in voting early (enough to risk such a punishment). Second, this incentive appears socially inefficient (assuming that the national parties' decisions reflect an interest in the efficiency of the whole nomination process).

In order to evaluate these arguments, consider the following – only half-fictional – example of a nomination contest with initially (at the time of the first elections) three serious contenders, whom we call C, E and O. These candidates differ on some characteristics that are relevant for voters. First, candidate C has experience in Washington and would know on day 1 where the light switches are in the White House, while candidate E and O run as "Washington outsiders" or "change candidates".[3] Suppose that some voters (ceteris paribus) prefer a candidate with Washington experience, while others (the "change voters") prefer an outsider. In addition, there is uncertainty about the "valence" (e.g. absence of scandals, proneness of gaffes etc.) of each candidate. Suppose first (and counterfactually) that the primary elections in all states occur simultaneously. In this case, it is quite plausible that C wins most states, as E and O split the change voters.

In contrast, in a sequential system, change voters in states that hold their primaries after the first ones can observe the early election results and coordinate their votes accordingly. For example, if O

[1] Different states have their presidential nomination elections organized as either primaries or caucuses. Since we are only interested in the temporal organization of the entire nomination process, we will, in a slight abuse of terminology, call all elections "primaries".

[2] Throughout the primary process, the Democratic National committee even threatened to take away all of Florida's and Michigan's delegates, but then reduced the size of the penalty to one-half.

[3] Of course, in reality, there were some differences between Edwards and Obama, but experience vs. change appears to have been the major fault line for the electorate in the early elections, and Edwards and Obama on the one side of this issue contrasted with Hillary Clinton on the other side. Moulitsas (2008b) documents convincing evidence from opinion polls that most Edwards and Obama supporters in Iowa and other states had the other change candidate as their second choice. For example, in a 12/26-30, 2007 poll by Opinion Research Corp for CNN, 36% of Iowa Democrats polled state that Edwards was their second choice, 25% name Obama, but only 11% name Clinton as their second choice. Since all three candidates were very close in terms of *first* preferences, this suggests that most Obama and Edwards supporters had the respective other candidate as their second preference.

gets more votes than E in the early elections, then even voters with ranking $E > O > C$ may vote for O, because they have determined that E has no chance of winning, and among the remaining relevant candidates, they prefer O. In this case, O will win the nomination if a majority of the electorate prefers him to C. For example, Moulitsas (2008a) cites a Rasmussen poll for Missouri from 1/31 (the last one conducted with Edwards in the mix) before the primary one week later. The preference numbers in the Rasmussen poll were Clinton 47, Obama 38, Edwards 11, while the actual election results were Obama 49.3, Clinton 47.1, Edwards 1.7. These numbers suggest that Edwards supporters migrated to Obama, after Edwards dropped out of the race.

The benefit of a sequential system is thus that the change voters are not split in most districts, thus making it more likely that a change candidate wins (if this is the preference of a majority of voters). There is, however, a cost: Conditioning coordination on only one or few initial elections raises the possibility that the weaker change candidate comes out on top, and if such an early electoral mistake would occur, it cannot be corrected in later districts precisely because of coordination. The objective of our model is to provide a formal framework for the analysis of the trade-off between coordination and information aggregation.

There is a widespread notion among political commentators and practitioners that candidates who do well in early primaries experience "momentum", i.e., are more likely to win in later primaries. Our model explains this notion as the result of voters in later primaries using the result of early primaries as focal points for coordination. In this sense, early primaries are particularly important, and candidates can therefore be expected to pay particular attention to these early contests. For example, Malbin (1985) reports that in the 1980 Republican primaries, George Bush and Ronald Reagan allocated roughly 3/4 of their respective total campaign budgets to states that held their primaries before March 31, although these states accounted for less than a fifth of the delegates to the Republican convention in 1980.

Many commentators and politicians have argued that the temporal organization of primaries may matter for the outcome, and some argue in favor of a sequential system, while others prefer a national primary. In our framework, we can analyze (i) under which circumstances the temporal organization makes a difference for who wins the nomination, and (ii) whether such a change is beneficial (from an ex-ante or utilitarian perspective). These points are not merely of theoretical interest, but also potentially policy-relevant because a reform of the primary system is frequently discussed.

Finally, there is a notion that states have an inefficient incentive to choose an early primary date, thus effectively making the primary system more like a simultaneous system. To generate such an incentive, the timing of elections must matter, and the states cannot be interested only in the quality of the selection process. We can easily extend our basic model to capture such incentives, and therefore can provide a rationale for the coordination role of national parties in the scheduling of primary dates.

The paper proceeds as follows. In the next section, we review related literature. In Section 3, we introduce our model. In Section 4, we analyze a special case of the model with three ideologically differentiated candidates and analyze the advantages and disadvantages of a sequential and a simultaneous primary system. The benefit of the special case is that it can be solved in closed form, but the drawback is that, by construction, the coordination benefit of sequential voting outweighs any potential benefits of simultaneous elections. A fairer comparison of the costs and benefits of both types of primary systems therefore requires the numeric analysis of the general model. Section 6 discusses extensions of the basic model to determine the optimal number of early primary states, the effects of the coordination role of primaries on candidates' campaign decisions and on the choice of states whether to hold their primaries early or late. Section 7 concludes.

2 Related Literature

Several political science studies analyze the relation between voters' expectations of which candidates will do well and their preference for these candidates. The study closest to our focus on the role of early primaries as a coordination device in Bartels (1987), which analyzes the 1984 Democratic presidential primary. Bartels describes the coordination process of those Democratic voters unhappy with the presumed front-runner, as follows.

> At the beginning of the 1984 primary season, the question facing prospective voters was whether or not to support the obvious front-runner, Walter Mondale. Those who were most predisposed to support Mondale(on the basis of issue preferences [...]) would do so without undue soul-searching. On the other hand, a fair number of Democrats who were lukewarm (or worse) about Mondale's candidacy may at least have entertained the possibility of supporting a different candidate. Their problem was to decide which alternative, if any, to turn to.
>
> Having framed the problem in this way, we may ask ourselves what a prospective voter with an eye out for an alternative to Mondale would have been likely to know about the other candidates in the race. At the beginning of the campaign, the best answer is probably 'very little'. But Hart's second-place finish in Iowa, followed by his dramatic upset victory in New Hampshire changed that. By the end of February, our prospective voter was quite likely to know at least one thing about at least one challenger: that Gary Hart was out there, an alternative to Mondale with significant popular support, [suggesting that] a vote for Hart would not be wasted.[4]

In the empirical part of the paper, Bartels does not focus on this coordination aspect (i.e., Hart versus other non-Mondale candidates), but rather analyzes the dynamic aspects of how expectations about the candidates' winning chances influenced voters' preferences.

Other studies analyzing similar relationships include Bartels (1985) (for the 1980 Democratic primaries) and Kenny and Rice (1994) for the 1988 Republican primary, but all of these focus implicitly

[4]Bartels (1987), pp. 13.

on a two-candidate framework. An exception to this is Knight and Schiff (2007), which provides both a theoretical model and an empirical study of the 2004 Democratic primary. In their model, voters in different states receive signals concerning the candidates' valences, as well as state-specific and individual preference shocks. Voters in later districts use the voting results in the earlier districts to infer the valence signal there. Then, voters vote for their highest-ranked candidate, given their current beliefs. Knight and Schiff use their model to empirically estimate the degree of social learning in the 2004 Democratic primaries with a discrete choice model and find substantial momentum effects are present. Voters in early primary states have a much higher influence on the selection of the nominee than voters in later primary states.

Our paper differs from Knight and Schiff (2007) in the route through which early election results influence later voters. In Knight and Schiff's model, voters in later primaries use the results of early primaries to infer the signals that voters in these states have received, and they use this inferred signal as well as their own private signal to update their prior beliefs about candidate quality. In contrast, we focus on the role of early primaries as a coordination device. The only influence of early election results is that they influence which candidates the voters in later districts consider to be relevant contenders. In practice, both learning and coordination may play important roles in the primary system, and so we view these approaches as complementary. Moreover, while Knight and Schiff (2007) is a positive model seeking to explain the role of momentum in sequential primaries, our central question is normative: Under which conditions is a sequential or a simultaneous primary system preferable from a social ex-ante point of view?

Most of the theoretical literature on primaries, reviewed below, has focused on elections with two alternatives (in which, naturally, the issue of coordination does not arise). An exception is Hummel (2008) who analyzes a model with three candidates, in which one half of the voters vote early and, after seeing those results, the other half vote. The winner is the candidate who receives the most votes overall. Voters know their preferences over candidates, but are uncertain about the preference distribution in the general population. Hummel shows that, while simultaneous elections often feature equilibria in which voters vote sincerely, this is not often the case in a sequential election system. In particular, in the late election, voters are likely to condition on the first period election result because it provides information about which candidates are likely to be the top contenders. A voter is much more likely to be pivotal between the two top contenders, which endogenously induces voters to focus on candidates who did well in the first round.

Dekel and Piccione (2000) analyze a model of sequential elections in which sophisticated voters try to aggregate their private information through voting. While, in principle, more information is available for voters in later elections, they show that every equilibrium of the simultaneous game is also an equilibrium of the sequential game, regardless of the sequence. The intuition for this result is that strategic voters know that their vote only matters if they are pivotal, and hence they behave as if they knew that all other voters are evenly divided between the two candidates. Thus, it does not matter for the election outcome which candidate is supported by the early voters.

While Dekel and Piccione (2000) show that the sequential structure does not allow voters to improve upon the information aggregation result that can be obtained with simultaneous elections, Ali and Kartik (2006) show that there are equilibria of the sequential game that do not correspond to equilibria of the simultaneous game. In particular, they construct an equilibrium in posterior-based voting in the context of a sequential election. In this equilibrium, if other voters play history dependent strategies, then it is individually optimal for each and every voter to do so as well. The information aggregation properties of such a herding equilibrium are worse than those of the equilibrium analyzed by Dekel and Piccione (2000). In summary, with two candidates, the design of a sequential primary system appears ill-advised from a social point of view, as the expected voting outcome is either the same or worse than in a simultaneous system.

Callander (2007) studies a sequential voting model which voters have an exogenous desire to vote for the winning candidate on top of common value preferences. Callander obtains equilibria which, at some point in the sequential election process, display bandwagon effects with certainty because the desire to conform eventually dominates information based voting.

Battaglini (2005) shows that, when voting is costly, the set of equilibria under simultaneous and sequential models are generically disjoint. In a related experimental paper, Battaglini, Morton, and Palfrey (2007) explore empirically the implications of voting costs in sequential and simultaneous elections.

Klumpp and Polborn (2006) analyze a contest model of sequential primaries in which two competing candidates have to choose how to allocate their campaign expenditures on the different states that hold their elections sequentially. In equilibrium, candidates allocate a large portion of their effort to the first states. There is momentum in the sense that the currently leading candidate has an increased equilibrium probability of winning the next election. From a normative perspective, they show that a sequential organization of primaries has the advantage of leading to lower expected expenditures than a simultaneous system.

In their seminal work on informational cascades, Bikhchandani, Hirshleifer, and Welch (1992) interpret "momentum" in primary races as evidence of herding on part of the later primary states. In a voting context, herding appears particularly inefficient: If all voters vote simultaneously (so that they don't have an opportunity to herd), the Condorcet jury theorem suggests that individual observation errors do not negatively affect the outcome, because the majority of voters will receive the correct signal. In contrast, if, in a sequential voting system, later voters condition their votes on the voters who are ahead of them in the sequence, then a lot of information may not be used, and the wrong candidate may get elected if he happened to be lucky and early voters received a wrong signal.[5] However, in this framework, it is difficult to understand why national parties have chosen to institute (and defend)

[5]It is unclear, however, whether primaries are really a valid example for herding: In herding models, people are concerned with *making the right choice*. In a standard model of voting, people do not so much care about whether *they themselves* voted for the correct candidate, but rather whether the best candidate is selected in the end.

a sequential primary system. In our model with three (or more) candidates, the act of using early elections in order to coordinate (which appears to be like a form of herding) may be socially beneficial, as it allows for coordination in situations where the plurality winner is not the Condorcet winner.

3 The model

There are K candidates who compete for their party's nomination as presidential candidate. Let $\mathcal{K} = \{1, \ldots, K\}$ denote the set of candidates, and let k denote a typical candidate. The set of electoral districts (i.e., states) is $\{1, \ldots, S\}$, with typical state s. We assume for simplicity that the number of states, S, is large and that all of them have the same size so that the candidate who wins the most districts wins the nomination (with ties broken by a coin flip). There are two election periods, and a district can either vote "early" (time 0) or "late" (time 1). Voters in late elections observe the election outcome in early elections.

Candidates differ in two dimensions. First, parameter v_k measures Candidate k's valence (which is a characteristic like competence appreciated by all voters). Second, there is a policy issue on which candidates have either position 0 or 1. Without loss of generality, we assume that the first k_0 candidates are fixed at $a_k = 0$, while the other $k_1 = K - k_0$ candidates are fixed at $a_k = 1$.

The policy dimension is meant to capture the idea that some candidates are very similar to each other and hence close (policy) substitutes for most voters, while there is a substantial difference to some other candidates. The assumption that policy differences can be expressed in binary form follows Krasa and Polborn (2007), and the assumption that there is only one major issue is just made for simplicity.

Voter i's utility from a victory of Candidate k is

$$U_k^i = v_k - \lambda |a_k - \theta^i| + \varepsilon_k^i. \tag{1}$$

Here, θ^i is Voter i's preferred position on the policy issue, and λ measures the weight of the policy issue relative to valence. The proportion of the total population in district s with preference for $a = 1$ is $\mu^s \in (0, 1)$, which is a random variable.

The last term, ε_j^i, drawn from $N(0, {\sigma_\varepsilon}^2)$ is an individual preference shock of voter i for Candidate j, as in probabilistic voting models. One possible interpretation of this term is that candidates also differ in a large number of other dimensions for which voters have different preferences. In this interpretation, the policy dimension modeled explicitly ($a_j = 0$ or $a_j = 1$) should be understood as the most important policy dimension.

There is uncertainty about the valence of the candidates. Specifically, each candidate's valence is an independent draw from a normal distribution $N(0, {\sigma_v}^2)$. Voters cannot observe v_j directly. Instead, voters in electoral district s observe a signal $Z_j^s = v_j + \eta_j^s$, where the additional term for Candidate j, η_j^s, is an independent draw from a normal distribution $N(0, {\sigma_\eta}^2)$. Note that η_j^s is a district-specific observation error (as opposed to a voter-specific observation error). The idea is that voters in the same state receive their news about the candidates from the same (local) news sources so that the errors (if any) are not individual-specific. Note that, if instead observation error terms were individual-specific, then the true valence of candidates would be known after the election results of the first district. This appears unrealistic.[6]

Given their own signal, and possibly the election results in earlier states (from which the signals in those earlier states can be inferred), voters rationally update their belief about the valence of candidates. Let $\hat{v}_j^s$ denote the valence of Candidate j that is expected by voters in district s.

As for voting behavior, we make the following assumptions. If district s is an early district (i.e., votes at time 0), then all of its voters vote sincerely. That is, voter i in district s votes for Candidate j if and only if

$$j \in \arg\max_{j' \in \mathcal{K}} \hat{v}_{j'}^s - \lambda |a_{j'} - \theta^i| + \varepsilon_{j'}^i. \tag{2}$$

(Since the distribution of ε is continuous, the measure of voters who are indifferent between 2 or more candidates is equal to zero, so it is irrelevant for the election outcome how those voters behave.)

Voters in "late states" (i.e. those voting at time 1) proceed as follows. From the election results of the early district(s), they determine a set $\mathcal{K}^1 \subset \mathcal{K}$ of *relevant candidates*, while the candidates in $\mathcal{K} \backslash \mathcal{K}^1$ are "eliminated". In particular, we focus on two rules rules for the determination of the set of relevant candidates which appear natural in that they focus on candidates that perform well in early elections. The *two-top-finisher rule* sets $\mathcal{K}^1 = \{$The two candidates who received the most votes in district 1$\}$. The *top-finisher by position rule* chooses the best finisher among the candidates with $a = 0$, and among the candidates with $a = 1$: $\mathcal{K}^1 = \{$Candidate 1, and the top finisher of Candidate 2 and 3 in the first district$\}$.[7]

Voter i in a late state s then votes for Candidate j if and only if

$$j \in \arg\max_{j' \in \mathcal{K}^1} \hat{v}_{j'}^s - \lambda |a_{j'} - \theta^i| + \varepsilon_{j'}^i. \tag{3}$$

Thus, the set $\mathcal{K}^1$ captures the notion that voters in later districts coordinate on those candidates who did well in earlier elections. Note that elections with three or more candidates usually have many Nash equilibria, even if voters eliminate weakly dominated strategies. Several papers provide arguments for why voters should be expected to coordinate on two of the candidates (e.g. Palfrey (1989), Messner and Polborn (2007), Hummel (2008)). As long as all voters believe that only some candidates are relevant (in the sense that the other voters only vote for these candidates), it is certainly an equilibrium for each voter to vote for his preferred candidate among the relevant candidates only.

Alternative coordination points to the two rules that we focus on appear less intuitive and often lead to lower ex-ante expected utility: For example, voters in late districts could focus on the weaker

[6]Of course, it appears plausible that, in reality, there is both a common as well as an idiosyncratic observation error. To simplify the model and gain some tractability, we focus on the state-specific observation error.

[7]Evidently, in many cases, these two rules produce the same set of relevant candidates, but sometimes, they may differ.

candidate in the first election, but this is neither intuitive nor useful from a the point of view of ex-ante utility. In another possible equilibrium, voters in late districts could ignore the early district and consider all candidates as relevant. In summary, it is important to stress that a sequential structure of primaries *facilitates* coordination (and the particular form of coordination that we focus on is, in our opinion, fairly natural), but, of course, sequential primaries do not *enforce* any particular form of coordination.

4 A simple model of coordination in primaries

In this section, we focus on a particular special case of the model that can be solved in closed form and helps us to gain some intuition for the effects of the temporal organization of primaries. We will analyze a more general version of the model in Section 5 below.

Here, we assume that there are three candidates ($K = 3$). Candidate 1's position is $a_1 = 0$, while the two other candidates have $a_2 = a_3 = 1$. Furthermore, we assume that λ is sufficiently large relative to the span of the distributions of valence υ such that a difference in the policy dimension (almost) always dominates both valence difference and the idiosyncratic preference shock ε. In other words, all voters with $\theta^i = 0$ will vote for Candidate 1, while those voters with $\theta^i = 1$ will either vote for Candidate 2 or 3. This creates a coordination problem for those voters whose preferred position is 1: If Candidates 2 and 3 split the votes of those voters who prefer position 1, then Candidate 1 may win even if he is not the Condorcet winner.

We also assume that the proportion of the total population with preference for $a = 1$ is equal to μ in all districts ($\mu^1 = \mu^2 = \ldots = \mu^N \equiv \mu$). Clearly, if $\mu < 1/2$, then Candidate 1 is the Condorcet winner, and his supporters form a majority in each district. If $\mu > 1/2$, then either Candidate 2 or Candidate 3 is the Condorcet winner, depending on which one of them has the higher valence.

We assume that the total number of states is very large ($S \to \infty$), and analyze two temporal organizations of the primary system. Under simultaneous elections, all S states vote at the same time (say, early). Under sequential elections, one state votes early, and the remaining $S - 1$ states vote late (after observing the election outcome in the first state).

4.1 Equilibrium

We now derive the equilibrium for the two different primary systems. It is useful to arrange the analysis by different values of μ. Remember that μ is the percentage of voters who have $\theta = 1$, i.e. prefer the policy position of Candidates 2 and 3. Our assumption that λ is sufficiently large implies that voters of type $\theta = 0$ vote for Candidate 1, while voters of type $\theta = 1$ vote for either Candidate 2 or 3.[8]

[8]In principle, the distribution of ε is unbounded such that there are some voters with, say, type $\theta = 1$, but a very large ε_1, who thus prefer Candidate 1. However, when λ is large relative to σ_ε, such voters will be exceedingly rare, and we will just ignore these cases in this section (in order to gain tractability).

Case 1: $\mu < 1/2$**.** Since $1 - \mu > 1/2$, Candidate 1 receives a majority of votes in every district. The election system only affects whether the votes of type $\theta = 1$ voters are split or united, but even coordination is insufficient for Candidate 2 or 3 to win in a single district.

Case 2: $1/2 < \mu < 2/3$**.** In this case, type 1 voters are in the majority, and given the preferences of voters, either Candidate 2 or 3 is the Condorcet winner in this polity. However, since Candidate 1 receives more than one-third of the votes, it is possible that he receives a plurality in some or all districts.

We first introduce some notation. Let ϕ_α, $\alpha \in \{v, \eta, \varepsilon\}$, denote the probability density function of the normal distribution of variable α. (Remember that all three variables are normally distributed with expected value 0, but different variances σ_α^2, $\alpha \in \{v, \eta, \varepsilon\}$). Furthermore, we denote by ϕ and Φ (without a subscript) the pdf and the cdf of the standard normal distribution $N(0, 1)$, respectively.

Consider first sequential elections. In the first district, Candidate 1 wins a percentage of $1 - \mu$ of the vote and thus either wins the first district, or finishes second. Thus, both the two-top-finisher rule and the top-finisher-by-position rule lead to the same set of relevant candidates in all later districts.

Since $\mu > 1/2$, either Candidate 2 and Candidate 3 (whoever wins more votes in the first district) will win all of the districts that vote late. Thus, in a sequential organization of primaries, it is impossible that Candidate 1 wins if he is the Condorcet loser.

However, the top-finisher among Candidate 2 and Candidate 3 in the first district is not necessarily the better candidate. Indeed, Candidate 2 gets more votes in the first district than Candidate 3 if and only if $v_2 + \eta_2^1 > v_3 + \eta_3^1$. Since $\eta_3 - \eta_2$ is distributed according to $N(0, 2\sigma_\eta^2)$, for given v_2 and v_3, Candidate 2 wins with probability $\Phi\left(\frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right)$. Note that $v_2 - v_3$ is distributed according to $N(0, 2\sigma_v^2)$. Without loss of generality, we can focus on the case $v_2 > v_3$; conditioning on this event, the density of $v_2 - v_3$ is given by $2\phi\left(\frac{t}{\sqrt{2}\sigma_v}\right)$. Thus, the probability that the better candidate wins is given by

$$2 \int_0^\infty \Phi\left(\frac{t}{\sqrt{2}\sigma_\eta}\right) \phi\left(\frac{t}{\sqrt{2}\sigma_v}\right) dt = \sqrt{2}\sigma_v \left[1 - \frac{\arctan\left(\frac{\sigma_\eta}{\sigma_v}\right)}{\pi}\right]. \tag{4}$$

Since the arc tan is an increasing function and lies between 0 and π (for positive arguments, such as here), it is easy to see that this probability is decreasing in σ_η and increasing in σ_v. Intuitively, a higher σ_η means that there is a larger chance that the difference of observation mistakes for the two candidates outweighs their valence difference, so that voters in the first district mistakenly perceive the worse candidate to be the better one. The probability in (4) is increasing in σ_v, as this increases the expected valence difference between the better and the worse candidate.

Consider now simultaneous elections. The voters in district s observe a signal $Z_j^s = v_j + \eta_j^s$. Using Bayes' rule, the updated expected value of Candidate j's valence is

$$\hat{v}_j^s = \int_{-\infty}^{\infty} \frac{\phi_v(t)\phi_\eta(Z_j^s - t)}{\int_{-\infty}^{\infty} \phi_v(t')\phi_\eta(Z_j^s - t')dt'} t dt = \frac{{\sigma_v}^2}{{\sigma_v}^2 + {\sigma_\eta}^2} Z_j^s. \tag{5}$$

If Voter i in district s has type $\theta = 1$, he votes for Candidate 2 if $\hat{v}_2^s + \varepsilon_2^i > \hat{v}_3^s + \varepsilon_3^i$, and for Candidate 3 if the inequality is reversed. Rearranging, the percentage of type $\theta = 1$ voters who vote for Candidate 2 is equal to

$$\text{Prob}(\varepsilon_3 - \varepsilon_2 \leq \hat{v}_2^s - \hat{v}_3^s) = \text{Prob}\left(\varepsilon_3 - \varepsilon_2 \leq \frac{{\sigma_v}^2}{{\sigma_v}^2 + {\sigma_\eta}^2}[Z_2^s - Z_3^s]\right) = \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right). \quad (6)$$

Similarly, Candidate 3's share of the vote of $\theta = 1$ types is equal to

$$1 - \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right).$$

Candidate 1, the Condorcet loser, receives all votes from $\theta = 0$ types (a proportion $1-\mu$ of the electorate) and wins a particular district s if and only if

$$1 - \mu > \mu \cdot \max\left(\Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right), 1 - \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right)\right), \quad (7)$$

hence if

$$\frac{2\mu - 1}{\mu} < \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right) < \frac{1-\mu}{\mu}. \quad (8)$$

Denoting the inverse of the cumulative distribution of the standard normal distribution by Φ^{-1}, and letting $\kappa = \frac{\sigma_v^2}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}$, we can write this as

$$\Phi^{-1}\left(\frac{2\mu - 1}{\mu}\right) < \kappa(v_2 - v_3) + \kappa(\eta_2 + \eta_3) < \Phi^{-1}\left(\frac{1-\mu}{\mu}\right) \quad (9)$$

For given v_2 and v_3, the term in the middle is normally distributed with expected value $\kappa(v_2 - v_3)$ and variance $2\kappa^2{\sigma_\eta}^2$. Thus, the percentage of districts won by Candidate 1 is given by

$$\begin{aligned}\text{Prob}\left(\Phi^{-1}\left(\frac{2\mu - 1}{\mu}\right) - \kappa(v_2 - v_3) < \kappa(\eta_2 - \eta_3) < \Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)\right) = \\ \Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) - \Phi\left(\frac{\Phi^{-1}\left(\frac{2\mu-1}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) = \\ \Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) - \Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right),\end{aligned} \quad (10)$$

where the last inequality uses the fact that $\Phi^{-1}\left(\frac{2\mu-1}{\mu}\right) = -\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)$, because $\frac{2\mu-1}{\mu}$ and $\frac{1-\mu}{\mu}$ are symmetric around 1/2 (i.e., add up to 1).

Again, suppose that $v_2 > v_3$, so that Candidate 2 is the toughest competitor for the nomination. The percentage of districts won by Candidate 2 is

$$\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right). \quad (11)$$

Candidate 1 wins the nomination if (10) is larger than (11) he wins more districts than Candidate 2, hence if

$$\Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)-\kappa(v_2-v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) > 2\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)-\kappa(v_2-v_3)}{\sqrt{2}\kappa\sigma_\eta}\right). \tag{12}$$

Note that the left hand side is decreasing in μ, while the right hand side is increasing in μ. Thus, if (12) holds for a particular level of μ, then it also holds for all smaller levels of μ (equivalently, all higher levels of $1-\mu$). This is intuitive, since $1-\mu$ is the percentage of voters who support Candidate 1. Let μ^* denote the level of μ such that (12) holds with equality.

Consider first the case of $\mu = 1/2$, such that $\frac{1-\mu}{\mu} = 1$ and hence $\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) = \infty$. Clearly, (12) holds, as the left hand side goes to 1, while the right hand side goes to 0. Intuitively, if $\mu = 1/2$, then any sort of vote-splitting between Candidates 2 and 3 guarantees that Candidate 1 wins all districts. Since both sides are continuous in μ, the same result holds (for any given v_2 and v_3) for μ sufficiently close to 1/2. Now consider the case of $\mu = 2/3$, such that $\frac{1-\mu}{\mu} = 1/2$. Since $\Phi^{-1}(1/2) = 0$, (12) is clearly violated.

Consider now the effect of changes in σ_ε, σ_η and σ_v on (12). Note first that $\kappa = \frac{\sigma_v^2}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2+{\sigma_\eta}^2)}$ is decreasing in σ_ε and increasing in σ_v. Furthermore, the left hand side of (12) is decreasing in κ (as $(1-\mu)/\mu > 1/2$, and thus $\Phi^{-1}(\frac{1-\mu}{\mu}) > 0$), while the right hand side is increasing in κ by the same argument. Thus, to preserve equality between the two sides of (12), an increase of κ needs to be balanced by a decrease of μ^*. Consequently, μ^* decreases in σ_v, and increases in σ_ε.

Intuitively, an increase of σ_ε means that voters with type $\theta = 1$ split their votes relatively evenly between Candidates 2 and 3, which makes it easier for Candidate 1 to win. In contrast, an increase in σ_v makes it harder for Candidate 1 to win, for two reasons. First, for fixed v_2 and v_3, κ increases. The reason is that a higher σ_v means that voters update more strongly after seeing their signals, and so, for fixed $v_2 - v_3$, the average voters' belief about $|v_2 - v_3|$ (average with respect to district signals) increases. Second, a higher σ_v also makes a higher valence difference between Candidate 2 and 3 more likely (i.e., increases the expected value of $|v_2 - v_3|$). A larger valence difference decreases the problem of vote splitting and makes it therefore more difficult for Candidate 1 to win.

The effect of σ_η on μ^* is ambiguous, as we show in the proof of Proposition 1 in the Appendix. To get an intuition for this result, consider the following two cases. First, suppose that $v_2 = v_3$ and that σ_η is very small; in this case, Candidates 2 and 3 split the votes of type 1 voters almost 50/50, so that μ^* is close to 2/3 (i.e., Candidate 1 is almsot always guaranteed to win for $\mu < 2/3$). An increase of σ_η leads to a less equal vote splitting and thus decreases μ^*. Second, suppose instead that $v_2 > v_3$, and that the difference is large relative to σ_ε. If σ_η is small, then almost all type 1 voters vote for Candidate 2, and therefore μ^* is close to 1/2. If, instead, σ_η is large, the estimated difference between v_2 and v_3 in the beliefs of voters is very likely quite small, which leads to substantial vote splitting and, thus, a high value of μ^*. Thus, there must be at least some parameter values for which and increase of σ_η leads to an increase of μ^*.

The worst case scenario for Candidates 2 and 3 in terms of vote splitting is if they are very similar in competence ($v_2 = v_3$). In this case, (12) becomes

$$\Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\sqrt{2}\kappa}\right) > 2\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\sqrt{2}\kappa}\right). \tag{13}$$

Since the normal distribution is symmetric, $\Phi(-x) = 1 - \Phi(x)$ for all x. Both sides of (13) become equal if $\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\sqrt{2}\kappa} = \Phi^{-1}\left(\frac{2}{3}\right)$. Solving for μ yields

$$\mu^* = \frac{1}{1 + \Phi\left(\sqrt{2}\kappa\Phi^{-1}\left(\frac{2}{3}\right)\right)}. \tag{14}$$

If $\mu > \mu^*$, then either Candidate 2 or Candidate 3 wins even if $v_2 = v_3$. Thus, a forteriori, if either of these candidates is strictly better than his competitor, he will win. For $\kappa \to 0$, μ^* goes to 2/3, and for $\kappa \to \infty$, μ^* goes to 1/2. Thus, if σ_v is small and/or σ_ε and σ_η are large, then μ^* goes to 2/3. The intuitive reasons for these relations are the same as discussed above.

Case 3: $\mu > 2/3$**.** In this case, Candidate 1 receives less than a third of the votes in every district, so that he loses in every district. Without loss of generality, suppose again that $v_2 > v_3$.

Under simultaneous elections, Candidate 2 wins in district s if

$$v_2 + \eta_2^s > v_3 + \eta_3^s. \tag{15}$$

Thus, for a given $v_2 > v_3$, the proportion of districts won by Candidate 2 is equal to $\Phi\left(\frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right) > 1/2$. Consequently, Candidate 2 is certain to win the nomination contest.

Under sequential elections, consider first a top-finisher by position rule (so that the winner of the first districts and Candidate 1 are the two relevant candidates in later elections). The winner of the first district gets a vote share μ in all following districts and thus wins the nomination. The probability that Candidate 2 is the winner of the first district is the same as in equation (4) in Case 2 above. Thus, the better candidate is likely to win the nomination, but there is a positive probability that the other candidate with the same policy position wins instead.

Now suppose instead that the set of relevant candidates is defined by the top-finisher rule. This rule makes a difference if Candidate 1 places last in the first district (which is possible if $\mu > 2/3$). In this case, the two candidates with policy $a = 1$ go on to compete in the remaining districts, and the stronger candidate wins a majority of these districts with probability 1, just like under simultaneous elections. Note, however, that $\mu > 2/3$ does not guarantee that Candidate 1 places last, and if he does not, then the outcome is exactly the same as under the top-finisher-by-position rule.

The following proposition sums up all of these results.

Proposition 1 *Suppose that voters utility function is given by* (1)*, and that λ is large relative to σ_v and σ_ε. Additionally, assume that Candidate 1's policy position is 0 and both Candidate 2 and 3 have policy position 1.*

1. *If $\mu < 1/2$, Candidate 1 is the Condorcet winner and wins the nomination under both a simultaneous and a sequential primary system.*

2. *If $1/2 < \mu < 2/3$, Candidate 1 is the Condorcet loser, and the candidate with the higher valence among Candidate 2 and 3 is the Condorcet winner.*
   In a simultaneous primary system, there exists $\mu^* \in (1/2, 2/3)$ given by (14) such that Candidate 1 wins the nomination with positive probability for every $\mu < \mu^*$, where μ^* is decreasing in σ_ε, increasing in σ_υ, and ambiguous in σ_η. The only other possible outcome in a simultaneous primary system is a victory of the Condorcet winner.
   In a sequential primary system, either Candidate 2 or Candidate 3 wins. The probability that the Condorcet winner wins is given by (4).

3. *If $\mu > 2/3$, Candidate 1 is the Condorcet loser, and the candidate with the higher valence among Candidate 2 and 3 is the Condorcet winner.*
   In a simultaneous primary system, the Condorcet winner wins with probability 1.
   In a sequential primary system, the Condorcet loser cannot win, but the Condorcet winner wins with probability strictly smaller than 1. Moreover, the top-2-finisher rule leads to a weakly better outcome than the top-finisher-by-position rule.

Proof. See Appendix. ■

5 Analysis of the general multi-candidate model

In the last section, we have assumed that valence and idiosyncratic preference shocks are relatively unimportant, so that type $\theta = 0$-voters would never vote for Candidate 2 or 3, and type $\theta = 1$-voters would never vote for Candidate 1. In this case, the only relevant uncertainty for voters is whether Candidate 2 or 3's valence is higher. Moreover, the two types of possible errors in Section 4 — electing Candidate 1 even though $\mu > 1/2$, and electing the less capable candidate among Candidate 2 and 3 — have very different welfare consequences. Since we assumed that the policy difference between Candidate 1 on the one side and Candidates 2 and 3 on the other side are much more important to voters than any valence difference, from a utilitarian (or ex-ante) perspective, the first type of error is much more severe than the second one.

If (what appears reasonable) the social planner chooses the election system at a time when the value of μ is unknown and drawn from some distribution, then, in the setup of Section 4, a sequential election system is optimal. More generally though, when the importance of policy questions does not necessarily swamp the importance of valence, there is a less trivial trade-off between the coordination benefits of a sequential system and the information aggregation advantage of a simultaneous system. As a welfare measure, it appears appropriate to consider a random citizen's expected utility under both types of primary systems.

We also want to expand the analysis to elections with more than just three candidates. The cost of going to a setting with more candidates (as well as one where policy difference between candidates do not always dominate the voters' candidate comparison) is that this model is too complicated to solve for the equilibrium in closed form. However, it is relatively straightforward to simulate numerically, which allows us to gain insights into how parameters influence the comparative static determinants of equilibrium and the optimality of the sequential or simultaneous voting systems.

5.1 Simulation procedure

All results reported below were obtained as follows. Each voter's utility ranking of candidates remains unaffected if we divide (1) by λ. Consequently, we can normalize $\lambda = 1$ without loss of generality. For all simulations, we set the number of states equal to $S = 49$. For a given set of remaining parameters $(K, \sigma_v, \sigma_\eta, \sigma_\varepsilon)$ and candidate fixed characteristics a_k, we first draw a vector of candidate valences $(v_1, \ldots, v_K)$ from $N(0, \sigma_v^2)$ and a general preference parameter μ_{gen} with the following interpretation: Of the 1001 voters in state s, $\mu_s = \mu_{gen} + M_s$ have ideal policy $\theta = 1$, where μ_{gen} is a general shock (equal for all states) drawn from a uniform distribution on $[450, 550]$, and M_s is a state-specific shock for state s, uniformly distributed on $[-100, 100]$. Thus, we allow for uncertainty about what is the preferred position of a majority in the whole economy, and we allow for the preference distribution to differ state-by-state. For each of the 49 states, we then draw values for η_s from $N(0, \sigma_\eta)$, and for each of 1001 voters per state we draw ε from $N(0, \sigma_\varepsilon)$.

We analyze simultaneous elections in which all states vote simultaneously, and sequential elections in which one state votes early and the remaining 48 ones vote late. In both electoral systems, voters update rationally about the candidates' valences, given the information that is available to them (i.e., their own signal, and, in case of a sequential election system, knowledge of the election result in the early state). Voters in simultaneous elections, and in the first district in a sequential election system, vote for the candidate who maximizes their respective expected utility given in (1). Voters in late districts in a sequential system vote for the candidate from among the set of relevant candidates who maximizes their respective expected utility.

For a given vector of candidate valences, we determine the outcome (i.e., the candidate who wins the nomination) and the average utility of voters under both the sequential and the simultaneous election system. We repeat this exercise 1000 times (i.e., we draw a new candidate valence vector and preference parameter μ_{gen}, and then proceed as described above), and report the average effects below.

Note that the number of voters per state (1001) is sufficiently large to ensure that the realized distribution of the idiosyncratic shocks is very close to the theoretical distribution (i.e., $N(0, \sigma_\varepsilon^2)$). As a consequence, the election outcome in state s is (almost) a deterministic function of the aggregate parameters: The realized values of v, η and μ_s, as well as σ_ε. Thus, increasing the number of voters per state would not significantly affect our results (but would, of course, increase computation costs proportionally).

Before we proceed to the results, it is useful to discuss the expected effects of parameters. The fundamental trade-off is between coordination, afforded by a sequential primary system, and information aggregation about valence, which works better in a simultaneous system. The parameters influence which of these effects is more important from an ex-ante point of view.

State specific shocks (σ_η). First, fix the number and position of candidates and consider the effects of state-specific shocks. In a simultaneous system, these shocks more-or-less cancel out (because the expected value of μ_s is equal to 50% of the electorate, and the expected value of η_s is equal to zero. The dispersion of the realized values of these random variables around their expectations should not significantly affect the overall performance of a simultaneous election system.

In contrast, in a sequential election system, a higher value of σ_η means that the first state is less likely to receive an approximately correct state-specific signal. Voter coordination on the election winner in the first state then becomes more problematic, as "herding" on the wrong candidate induced by the election outcome in the first state cannot be corrected in the remaining districts. Thus, we expect that a higher σ_η favors a simultaneous system.

Individual-specific shocks (σ_ε). Now consider the effect of the size of the individual-specific shocks ε_j^n. If σ_ε is very small, then almost all voters in each state who have the same type ($\theta = 0$ or $\theta = 1$) have the same preference ranking over candidates. Hence, under simultaneous elections, the district is very likely won by the candidate who has the policy position favored by the majority and who, among all such candidates, has the most favorable signal $Z_j^s = v_j + \eta_j^s$. While the signal may be wrong in a particular district, it is very likely to be correct in a majority of states. Hence, for very small σ_ε, simultaneous elections are likely to yield better expected outcomes.

In contrast, when σ_ε is large, then the candidate with the highest $Z_j^s = v_j + \eta_j^s$ among all candidates with the same position will still get a plurality of the votes from voters of the corresponding type, but a substantial fraction of these votes will go to other candidates.[9] Now, if there are fewer candidates with the minority policy position than those with the majority policy position, then it is possible and even likely that one of the minority candidates wins in sufficiently many states to secure the nomination. Thus, a higher σ_ε makes a simultaneous system less attractive.

Consider now a sequential system. The distribution of ε has no effect on which of the candidates with $a = 0$ and $a = 1$ ends up with the most votes in comparison to the other candidates with the same ideological position, respectively. Therefore, when σ_ε is large, then a sequential system is likely to be optimal.

Candidate valence. When σ_v is small, then all candidates have very likely approximately the same valence. In this case, the main determinant of the efficient election outcome is whether type 0 voters

[9] Note that this vote-splitting argument does not depend on any uncertainty about the aggregate distribution of ε (which, as argued above, is essentially deterministic, given the large number of voters).

or type 1 voters are in the majority. This is best determined by a sequential election system, because it avoids any coordination mistakes.

When, in addition, $\sigma_v/\sigma_\varepsilon$ is very small, then all $a = 0$ candidates receive approximately the same number of votes. The same holds for all $a = 1$ candidates. Thus, if $\mu_1/k_1 < (1 - \mu_1)/k_0$, i.e., the number of type 1 voters per type 1 candidate is smaller than the number of type 0 voters per type 0 candidate, then the top-two vote getters in the first district are likely both type 1 candidates. In this case, the coordination advantage is considerably higher if voters follow a top-finisher-by-position rule rather than a two-top-finisher rule.

In contrast, when σ_v is higher, the disadvantage of the sequential system becomes more severe, as voters may coordinate on the wrong candidate. To see this, observe that, when valence differences dominate policy differences, voters vote mostly based on (perceived) valence, and even in a three-way simultaneous race, the candidate with the highest valence is very likely to come out on top. Thus, for σ_v large, a simultaneous election system is likely preferable to a sequential one.

Candidate field composition. Intuitively, the benefit of coordination afforded by a sequential primary system depends on how severe the problem of vote splitting is, and how much more one group (of candidates with the same fixed policy) is affected from vote splitting relative to their ideological competitors.

For example, start from the setting in Section 4 where there are three candidates, two of whom have type $a = 1$. If we add a fourth candidate who has type $a = 0$, then both ideological camps are similarly affected from vote splitting, and the relative advantage in terms of expected utility of a simultaneous rather than sequential election system should increase. In contrast, if the fourth candidate has type $a = 1$ (so that there is now a 1/3 split among candidates, according to their policy), then vote splitting becomes an even more severe problem among the $a = 1$ candidates than in the initial situation, and consequently the attractiveness of a sequential voting system should increase.

Therefore, all other things equal, a simultaneous election system becomes more attractive with 5 candidates.

Importance of policy differences. Remember that we can divide (1) by λ without affecting any voter's preference ranking. This allows us to normalize the importance of policy, $\lambda = 1$ in our simulation. Effectively, an increase in λ (i.e., an increase in the importance of policy for voters) is the same as a simultaneous decrease of σ_v, σ_η and σ_ε, all by the same factor.

Intuitively, if policy differences are not at all important, then the competition between those candidates who share the same policy is only marginally more intense. Most voters make their voting decisions based on valence. The electoral system which is most likely to select the candidate with the highest valence in this scenario is therefore a simultaneous primary system.

In contrast, if policy differences are very important, then almost all type 0 voters vote for Candidate 1, while almost all type 1 voters vote for either Candidate 2 or 3. Moreover, it is desirable that either Candidate 2 or 3 wins if and only if type 1 voters are in the majority. This can be achieved in a sequential primary system, while in a simultaneous system, vote splitting between Candidates 2 and 3 likely means that Candidate 1 wins, even if type 0 voters are a minority.

The following Table 1 summarizes the comparative static predictions discussed so far.

Parameter change	Which system becomes more attractive?
$\sigma_\eta \uparrow$	simultaneous
$\sigma_\varepsilon \uparrow$	sequential
$\sigma_v \uparrow$	simultaneous
$\lambda \uparrow$	sequential
Candidate distribution more lopsided	sequential

Table 1: Comparative static predictions

5.2 Results.

We ran simulations for all 12 combinations of $\sigma_\varepsilon \in \left\{\frac{1}{4}, \frac{1}{2}\right\}$, $\sigma_\eta \in \left\{\frac{1}{5}, \frac{2}{5}\right\}$ and $\sigma_v \in \left\{\frac{1}{3}, \frac{2}{3}, 1\right\}$

For example, consider the case of $\sigma_\varepsilon = 1/4$, $\sigma_\eta = 1/5$ and $\sigma_v = 2/3$ (see Table 2). In a simultaneous system, one of the two candidates with $a = 0$ won in 496 out of 1000 runs, while in the remaining 504 cases, one of the three candidates with $a = 1$ won. In a sequential system, instead, the $a = 0$ candidates won a combined number of 411 runs. Thus, as expected, a simultaneous system benefits the candidates who suffer less from vote splitting (here, the two candidates with $a = 0$).

	Sequential elections relative to simultaneous elections
Change of nominee	218
victory of an $a = 1$-candidate (change)	589(+85)
...increased/decreased valence	115/103
...increased/decreased policy support	138/48
Δ ex-ante utility	+0.0084
...as percent of difference between simultaneous elections and SP solution	+15%

Table 2: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$, $\sigma_v = 2/3$

In 218 out of 1000 runs, the organization of the primary system (i.e., simultaneous or sequential) mattered for the identity of the election winner. In 115 of these cases, the valence of the overall winner was higher under simultaneous elections, while in the remaining 103 cases, the valence of the winner was higher under sequential elections. Thus, for these parameter values, there was a slight advantage for the simultaneous system in terms of selecting the candidate with the highest valence.

The other desirable feature of an election system is to select a candidate whose position is favored by a majority of voters. Remember that majority preference is a random variable in our simulations: The average percentage of voters with a preference for $a = 0$, drawn anew for each run, is uniformly distributed between 45 and 55. In 138 runs (out of the 218 in which the election system mattered), the total number of voters who preferred the winner's position was higher under sequential elections. In 48 runs, the opposite relation held. The remaining 32 runs are cases where the identity of the winner changed between two candidates with the same position (e.g., under simultaneous elections, Candidate 4 wins, while under sequential elections, Candidate 5 wins; both of these candidates have $a = 1$. In summary, a sequential election system has a significant advantage (for these parameter values) with respect to the selection of a candidate whose policy position is preferred by a majority of voters.

Overall, a voter's ex-ante expected utility is slightly higher under a sequential election system (by 0.0084). To gauge the size of this effect, consider a hypothetical social planner who maximizes a utilitarian welfare function. A voter's ex-ante expected utility with the social planner is about 0.0556 higher than under a simultaneous election system. Thus, the improvement in ex-ante expected utility brought from going from a simultaneous to a sequential election system is about 15 percent of the maximum possible welfare gain relative to a simultaneous election system.

Consider now how parameter changes affect these results. For $\sigma_v = 1/3$ (see Table 3 in the Appendix), the difference between the election outcome in simultaneous and sequential elections is more pronounced: 371 of the 1000 races are affected. Since valence is less likely to make a significant difference for voters, voters split mainly along ideological lines. In a simultaneous primary system, this means that the two candidates with $a = 0$ have a much better chance to win, as there is less vote splitting for them than for the three candidates with $a = 1$. Specifically, in a simultaneous system, only 338 campaigns are won by a candidate with $a = 1$. Under sequential elections, this number increases by 196. Again, as in the baseline case, a simultaneous primary system chooses more often a candidate with the higher valence than a sequential system, while a sequential system in better in selecting a candidate with majority ideological support. These two relationships hold for any of the parameter combinations analyzed and reported in Tables 2 to 13. In terms of welfare, for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$ and $\sigma_v = 1/3$, ex-ante utility is higher under a sequential system.

With the same parameters but $\sigma_v = 1$ (see Table 4), the primary system is less likely to make a difference; the nominee changes only in 131 runs. Even under simultaneous elections, all candidates have ex-ante approximately equal chances of winning. Consistent with our prediction that a higher σ_v should make a simultaneous system more attractive, ex-ante expected utility is higher under a simultaneous system, but the difference is quantitatively small.

Returning to the baseline case except for setting $\sigma_\eta = 0.4$ (see Table 6) yields a more substantial advantage of simultaneous elections in terms of expected utility. Intuitively, when σ_η is high, the first district top performers are less likely to be really the best candidates, and in a sequential system, such an error cannot be corrected any more.

The results for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.2$ and $\sigma_\upsilon = 2/3$ (i.e., a higher σ_ε than in the benchmark case) are reported in Table 9. Consistent with the expectation, a higher σ_ε increases the relative benefit of a sequential primary system.

Remember that an increase in λ is equivalent with a proportional decrease of $(\sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon)$. For example, we can compare the case that $(\sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon) = (0.5, 0.4, 2/3)$ (in Table 12) with the case that $(\sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon) = (0.25, 0.2, 1/3)$ (in Table 3). As expected, the increase in λ (decrease in the standard deviations) makes sequential elections more attractive, with ex-ante expected utility increasing from -0.04 to $+0.004$. Note that, in the case of $(\sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon) = (0.25, 0.2, 1/3)$, the election system more often makes a difference for the election outcome (as valence effects are less likely to dominate), and changes are more likely to be beneficial.

The largest absolute advantage for a simultaneous primary system relative to a sequential one arises for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.4$ and $\sigma_\upsilon = 2/3$. For these parameters, the ex-ante utility difference between the two systems is about 0.04. The key to this result is the high value of σ_η which makes an "error" of the first district more likely.

Note that, when $\sigma_\varepsilon = 0.25$ instead of 0.5 (while $\sigma_\eta = 0.4$ and $\sigma_\upsilon = 2/3$), the advantage of a simultaneous system in terms of expected utility is smaller (0.027 rather than 0.04). This is surprising in the sense that a lower σ_ε makes vote splitting less severe and therefore (as we argued above) should make it more likely that the optimal system is simultaneous. However, if we start in a situation where the optimal system is simultaneous, then a decrease in σ_ε need not increase the *absolute* advantage of the simultaneous system. Intuitively, suppose that σ_η is large (but finite), and consider what happens when σ_ε goes to infinity. When all candidates are relevant, Candidates 1 and 2's vote totals are almost identical, as are the vote totals of Candidates 3, 4 and 5. Since Candidates 1 and 2 have some advantage in that they have a larger pool of voters per candidate who prefer their policy position than the other three candidates, they are likely to be the top performers in each district under simultaneous primaries, and in the first district under sequential primaries. Thus, the outcome in both systems is very likely a victory of the higher-valence candidate among Candidates 1 and 2. In terms of ex-ante expected utility, the systems deliver approximately similar results.

Now consider what happens when the value of σ_ε is intermediate. In simultaneous elections, Candidates 3, 4 and 5 are likely to split the votes of $\theta = 1$ voters and therefore either Candidate 1 or 2 will win in each district. The overall result is the same as above, with a victory of the higher-valence candidate among Candidates 1 and 2 as the most likely outcome. Under sequential elections, instead, one of the three $a = 1$ candidates is likely to proceed as a relevant candidate after the first election, and will win if a majority of voters prefers $a = 1$. However, if σ_η is large (relative to σ_υ), then this successful candidate who proceeds is not necessarily the one with the highest valence among Candidates 3, 4 and 5. Thus, switching from simultaneous to sequential elections may replace the higher valence candidate among Candidate 1 and 2, with a more-or-less random draw from among Candidates 3, 4 and 5. When, in addition, the electorate is pretty evenly split, the advantage of choosing a candidate with the majority-preferred position is not that large, and a simultaneous system thus can have a substantial advantage over a sequential one.

Finally, consider Table 14, which reports the results for the baseline case of $(\sigma_\varepsilon, \sigma_\eta, \sigma_v) = (0.25, 0.2, 2/3)$, but assumes that there is only one candidate whose position is $a = 0$ and two whose position is $a = 1$. In this case, we expected that the asymmetry is more severe than with five candidates split 2/3. The nominees differ between sequential and simultaneous elections 249 times (and hence more often than in the baseline case with 5 candidates, where the election system mattered in 218 runs). In those cases where the election system matters, the effects are more likely to be positive for the sequential election system than in the baseline case. In particular, in a majority of cases (132 vs. 117 runs), the valence is now higher with the sequential system.

Overall, a voter's ex-ante expected utility is higher under a sequential election system (by 0.01, about one sixth more than in the baseline case). This improvement in ex-ante expected utility brought from going from a simultaneous to a sequential election system is about 15 percent of the maximum possible welfare gain relative to a simultaneous election system.

6 Extensions

We now discuss several extensions of our model. For the sake of simplicity, we sometimes refer in our discussion to the case of three candidates (as in Section 4), but it should be clear that none of the basic insights depends on this particular setup.

More states in first period So far, we have assumed that, in a sequential system, one state goes first and then all remaining states move simultaneously. The Condorcet jury theorem suggests that it may be better to have more than one election early.

In fact, if the distribution of the state-specific error η is independent of the temporal organization of primaries, then an optimal organization is easy to construct. In the first time period, there should be contests in many districts, though these should number fewer than half of the total number of districts. This way, the probability that the better one of Candidate 2 and 3 wins more votes than his competitor with the same ideological position is maximized (and, if S is large, is close to 1). This candidate then becomes the relevant candidate for the late elections, and wins all of these elections whenever $\mu > 1/2$. This is sufficient to secure the nomination even if Candidate 1 had won all of the early elections.

This argument is correct as long as the probability of a correct signal does not depend on the number of states that vote early. However, in the election context, it appears more plausible that the signal quality in a state depends on the extent to which candidates campaign in a state. If only one state votes early, candidates will campaign heavily in that state, and the probability that voters are well informed is relatively high. In contrast, the more states vote early, the fewer time do the candidates have to campaign in each state, and so it appears plausible that the probability of a correct state-specific signal in an early election state is a decreasing function of the number of states that hold their election early. In such a setup, it may well be the case that the optimal sequential primary system features only very few states that vote early.

Candidate actions. Candidates do not take any actions in our model, but it would be straightforward to incorporate this in our model. Suppose, for example, that candidates can take an action that increases their winning probability in a district, but is also costly (either directly, or by decreasing the winning probability in other districts). One can think of this action either as heavy campaign spending, or as pandering to special interests of the target district, which may be unpopular in other districts.

In a simultaneous primary system, candidates have very limited incentives to treat any one of the states in a very particular fashion. Assuming (which appears reasonable) that there are decreasing returns to campaigning and pandering, candidates will distribute their campaign efforts more or less uniformly over all states.

In contrast, in a sequential primary system, there is a very strong incentive for Candidates 2 and 3 to focus their campaign resources and/or electoral promises on the early state. The reason is that each of them has to win this state, or else be knocked off by the other candidate. Only one of them will go on to battle Candidate 1 in the remaining states.

Our coordination theory of primaries is consistent with the observation that (many) candidates spend a disproportionate amount of resources in early contests that award few delegates for the national conventions. As an extreme example in which (some) candidates spend a lot of resources even though no delegates are at stake, consider the Iowa Straw Poll in August of years before presidential elections. “As a straw poll, the Ames Straw Poll’s results are non-binding and have no official effect on the presidential primaries. However, the straw poll is frequently seen as a first test of organizational strength in Iowa by the news media and party insiders. As such, it can become very beneficial for the winning candidate on the national level because it builds momentum for their campaign.”[10] However, while some candidates choose to “participate” (i.e., often sponsor transportation and the $ 35 ticket costs for their supporters), other candidates bypass the event. In 2007, Mitt Romney won the Iowa Straw Poll, but the contest between Mike Huckabee and Sam Brownback was the most significant result for the development of the national campaign. Huckabee and Brownback competed for the support of a very similar demographic, evangelical Christians. Huckabee received 18.7% in the Iowa Straw Poll, versus Brownback’s 15.3%. “Fundraising and visits to [Brownback’s] website declined dramatically after this event, as many supporters had predicted Brownback would do much better, and speculation began that the candidate was considering withdrawing from the campaign.”[11] Brownback officially ended his campaign two months later. Huckabee consolidated the support of evangelicals and proceeded to win the Iowa caucuses in January 2008.

In contrast to the two candidates with the same fixed policy, Candidate 1 in our model should not use too much of his resources in state 1: Since there is no coordination issue in which Candidate 1 is

[10] Wikipedia, http://en.wikipedia.org/wiki/Ames_Straw_Poll.

[11] http://en.wikipedia.org/wiki/Sam_Brownback

involved, the first state is just one of the many that are needed for an overall victory. This is consistent with observations from the 2008 Democratic nomination contest. Both "change" candidates, Barack Obama and John Edwards spent a very significant proportion of their time and resources campaigning in Iowa, while Hillary Clinton focused much more on a national campaign. In the Iowa Caucuses, Obama won with 37.6% of delegates, Edwards came in second with 29.8% and Clinton third with 29.5%. However, while Edwards beat Clinton, the main result was that he lost to Obama, and the "change" voters in the following states would focus on Obama (especially after Obama also got more votes than Edwards in New Hampshire, the next contest). At the end of January 2008, Edwards suspended his campaign. In contrast, Hillary Clinton did not suffer that much from her third place finish in Iowa, because the voters attracted to her did not have any close (and better-performing) substitute to vote for instead.

Incentive for states to choose early primaries. In the U.S. presidential primary system, states appear to have an incentive to hold their primaries early, while the national parties would prefer most states to go for a late date. For example, in the 2008 presidential primaries, the earliest primary date that the national committees of both major parties allowed for states in general was February 5th, and indeed, most states chose to hold their primaries *exactly* on February 5th, indicating that this constraint was binding. Only Iowa, New Hampshire and South Carolina (as well as Nevada for the Democrats) were allowed to hold their nomination contests earlier.

Without the consent of the parties, Michigan and Florida chose to hold their primaries in January. As a consequence, the Republican and Democratic National Committee decided to take away half or all of the delegates of these two states to the nomination convention, respectively. The Democrats also offered to award additional delegates to states that vote late. Thus, national parties display a preference for states to vote truly sequentially, while individual states, as exemplified by Michigan and Florida, but also the other states that chose their first allowed primary date, appear to have a preference for being voting as one of the earliest states. In this context, it is also interesting to note that New Hampshire entitles its secretary of state to unilaterally reschedule its primary date, if this becomes necessary to preserve its status as the first-in-the-nation primary.

This tension between the national parties and single states is discussed in Cost (2007). He cites from an article in the Detroit Free Press: " 'When they're making promises ... we want to make sure they're not just making promises to Iowa, New Hampshire and South Carolina,' said state Senate Minority Leader Mark Schauer, D-Battle Creek." Cost argues that states play a prisoners' dilemma type game when scheduling their primaries, and that the national parties have the task of coordinating the states' dates and enforcing the socially efficient allocation.

Our model can easily be extended to explicitly model such incentives. As described in the last subsection, assume that candidates can choose for each district, one policy that "panders" to that district's voters, i.e. a policy that the voters in the target district like and everybody else dislikes. The candidates for whom the first district is decisive for reasons of voter coordination in later districts will actively pander there, even if this decreases their appeal in later districts.

In contrast, if all elections occur simultaneously, then such pandering is not very attractive for candidates as the votes won in the target district are counterbalanced by votes lost in other districts. Consequently, districts that vote at the same time as many other districts (either in a simultaneous primary system, or as late states in a sequential system), are unlikely to receive much pandering by candidates. Thus, in a sequential system, states have an incentive to move from the class of late states (i.e., those states in which voters coordinate based on the candidates' previous performance) to the class of early states.

Thus, there are two possible reasons why a decentralized choice of primary dates by individual states may lead to an inefficient primary schedule. First, given that candidates have an incentive to pander to early states, we would like to keep the total extent of pandering as small as possible, and thus it would be desirable to have only very few early states. Also, it is desirable to choose small states for this role, as conceivably the overall cost of pandering to voters in other states is positively related to the size of the state that candidates pander to.

Second, when all, or a sufficiently large majority of, states choose to hold their primary early, then the coordination benefit of sequential primaries is lost. Individual states may all choose to go early even if a sequential organization may be better for the expected utility of all of them. Both reasons provide a justification for the centralized coordination of primary dates managed by national parties.

7 Conclusion

At the beginning of presidential primaries, there are often several serious contenders. Some of them may be ideologically close substitutes for voters, while the difference to other candidates may be more significant. In a simultaneous election with a large set of candidates, the candidate who would come out on top is not necessarily the Condorcet winner. In contrast, sequential elections allow voters to narrow down the field of contenders as a way of avoiding vote-splitting among ideologically similar candidates. The sequential nature of the primaries therefore likely has facilitated the victory of candidates who were not the frontrunner at the beginning of the primary season, such as Obama (and possibly McCain) in 2008, and the very strong showing of Gary Hart in 1984.

In this paper, we have presented a model of voting in sequential primaries based on the idea of coordination. We assume that voters in late primaries consider only those candidates who performed well in early contests to be relevant candidates, and they vote for their favorite candidate from this restricted set. Thus, in contrast to simultaneous elections, sequential elections allow for voters to coordinate.

We show that, from an ex-ante perspective, this may be beneficial or detrimental. While sequential elections have the advantage of allowing voters to coordinate (and thus avoid that a candidate wins just because his ideological opponents split the votes of their supporters among each other), the disadvantage of sequential elections is that, once coordination has occurred, there is no possibility to correct an error made in early elections.

We show that sequential elections are likely to dominate simultaneous ones if valence differences between candidates are small; if the signal quality in early states is high; if there is a lot of vote-splitting between ideologically similar candidates; and if the ideological composition of the candidate field is asymmetric. In contrast, when valence differences are important, vote-splitting is not too important and the signal quality is bad, then a simultaneous primary system is superior.

References

Ali, N. and N. Kartik (2006). Social learning in elections. UC San Diego.

Bartels, L. M. (1985). Expectations and preferences in presidential nominating campaigns. *The American Political Science Review 79*(3), 804–815.

Bartels, L. M. (1987). Candidate choice and the dynamics of the presidential nominating process. *American Journal of Political Science 31*(1), 1–30.

Battaglini, M. (2005). Sequential voting with abstention. *Games and Economic Behavior 51*, 445–463.

Battaglini, M., R. Morton, and T. Palfrey (2007). Efficiency, equity and timing in voting mechanisms. *American Political Science Review 101*, 409–424.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*, 992–1026.

Callander, S. (2007). Bandwagons and momentum in sequential voting. *Review of Economic Studies 74*, 653–684.

Cost, J. (2007). Michigan, too? Horseraceblog on realclearpolitics.com, August 31, 2007; available at http://www.realclearpolitics.com/horseraceblog/2007/08/michigan_too.html.

Dekel, E. and M. Piccione (2000). Sequential voting procedures in symmetric binary elections. *Journal of Political Economy 108*(1), 34–55.

Hummel, P. (2008). Sequential voting in large elections with multiple candidates. mimeo, Stanford.

Kenny, P. J. and T. W. Rice (1994). The psychology of political momentum. *Political Research Quarterly 47*(4), 923–938.

Klumpp, T. and M. K. Polborn (2006). Primaries and the new hampshire effect. *Journal of Public Economics 90*, 1073–1114.

Knight, B. and N. Schiff (2007). Momentum and social learning in presidential primaries. NBER.

Krasa, S. and M. K. Polborn (2007). The binary policy model. University of Illinois.

Malbin, M. (1985). You get what you pay for, but is that what you want? In G. Grassmuck (Ed.), *Before Nomination: Our Primary Problems*. American Enterprise Institute for Public Policy Research.

Messner, M. and M. K. Polborn (2007). Strong and coalition-proof political equilibria under plurality and runoff rule. *International Journal of Game Theory 35*, 287–314.

Moulitsas, M. (2008a). Edwards voters' second choice, part II. Daily Kos, http://www.dailykos.com/storyonly/2008/8/12/1101/20327/19/566444.

Moulitsas, M. (2008b). Wolfson's ridiculous speculation. Daily Kos, http://www.dailykos.com/storyonly/2008/8/11/163149/067/156/566331.

Palfrey, T. R. (1989). A mathematical proof of Duverger's law. In P. Ordeshook (Ed.), *Essays in Contemporary Political Theory*. University of Michigan Press.

8 Appendix

8.1 Proof of Proposition 1

Proof. Most arguments are in the text before the proposition. Here, we analyze the effect of σ_η. Consider the difference of the left-hand and right-hand side of (12), and substitute for κ and set the expression equal to 0 (which implicitly determines the value of μ^*); this yields

$$Z = \Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right) - 2\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right) = 0. \tag{16}$$

Since $\Phi(\cdot)$ is an increasing function, $\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)$ is decreasing in μ, and thus $\frac{\partial Z}{\partial \mu}$. Consequently, the sign of

$$\frac{d\mu^*}{\sigma_\eta} = -\frac{\frac{\partial Z}{\partial \sigma_\eta}}{\frac{\partial Z}{\partial \mu}}$$

is the same as the sign of $\frac{\partial Z}{\partial \sigma_\eta}$. We have

$$\begin{aligned}\frac{\partial Z}{\partial \sigma_\eta} = &\left[\phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right) + 2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right)\right] \times \left[\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)\frac{\sigma_\eta^2 - \sigma_v^2}{\sigma_\eta^2}\right] + \\ &\left[\phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right) - 2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right)\right] \times \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta^2}\end{aligned} \tag{17}$$

(17) is greater than

$$2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right)\left[\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)\frac{\sigma_\eta^2 - \sigma_v^2}{\sigma_\eta^2} - \frac{v_2 - v_3}{\sqrt{2}\sigma_\eta^2}\right]$$

Since the term in square brackets goes to $\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) > 0$ for $\sigma_\eta \to \infty$, (17) is positive for σ_η sufficiently large. Thus, for σ_η sufficiently large, $\frac{d\mu^*}{d\sigma_\eta}$ is positive.

In contrast, for $v_2 = v_3$ and $\sigma_\eta < \sigma_v$, (17) and hence $\frac{d\mu^*}{d\sigma_\eta}$ is negative. ■

8.2 Simulation results

We report here the results of the simulations for all 12 combinations of $\sigma_\varepsilon \in \left\{\frac{1}{4}, \frac{1}{2}\right\}$, $\sigma_\eta \in \left\{\frac{1}{5}, \frac{2}{5}\right\}$ and $\sigma_v \in \left\{\frac{1}{3}, \frac{2}{3}, 1\right\}$ (except for the case of $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$, $\sigma_v = 2/3$, which was reported in Table 2 in the main text).

	Sequential elections relative to simultaneous elections
Change of nominee	371
victory of an $a = 1$-candidate (change)	534 (+196)
…increased/decreased valence	163/208
…increased/decreased policy support	222/84
Δ ex-ante utility	+0.0040
…as percent of difference between simultaneous elections and SP solution	+5%

Table 3: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$, $\sigma_\upsilon = 1/3$

	Sequential elections relative to simultaneous elections
Change of nominee	131
victory of an $a = 1$-candidate (change)	601 (+31)
…increased/decreased valence	49/82
…increased/decreased policy support	86/24
Δ ex-ante utility	-0.00032
…as percent of difference between simultaneous elections and SP solution	-1%

Table 4: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$, $\sigma_\upsilon = 1$

	Sequential elections relative to simultaneous elections
Change of nominee	484
victory of an $a = 1$-candidate (change)	516 (+293)
…increased/decreased valence	191/293
…increased/decreased policy support	273/93
Δ ex-ante utility	-0.025
…as percent of difference between simultaneous elections and SP solution	-20%

Table 5: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.4$, $\sigma_\upsilon = 1/3$

	Sequential elections relative to simultaneous elections
Change of nominee	308
victory of an $a = 1$-candidate (change)	592 (+107)
...increased/decreased valence	120/188
...increased/decreased policy support	144/79
Δ ex-ante utility	-0.027
...as percent of difference between simultaneous elections and SP solution	-47%

Table 6: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.4$, $\sigma_v = 2/3$

	Sequential elections relative to simultaneous elections
Change of nominee	196
victory of an $a = 1$-candidate (change)	593(+39)
...increased/decreased valence	57/139
...increased/decreased policy support	92/45
Δ ex-ante utility	-0.035
...as percent of difference between simultaneous elections and SP solution	-144%

Table 7: 1000 simulated campaigns for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.4$, $\sigma_v = 1$

	Sequential elections relative to simultaneous elections
Change of nominee	344
victory of an $a = 1$-candidate (change)	546(+140)
...increased/decreased valence	181/163
...increased/decreased policy support	179/98
Δ ex-ante utility	0.01
...as percent of difference between simultaneous elections and SP solution	13%

Table 8: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.2$, $\sigma_v = 1/3$

	Sequential elections relative to simultaneous elections
Change of nominee	180
victory of an $a = 1$-candidate (change)	605(+67)
. . . increased/decreased valence	105/75
. . . increased/decreased policy support	91/58
Δ ex-ante utility	0.014
. . . as percent of difference between simultaneous elections and SP solution	35%

Table 9: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.2$, $\sigma_v = 2/3$

	Sequential elections relative to simultaneous elections
Change of nominee	97
victory of an $a = 1$-candidate (change)	609(+27)
. . . increased/decreased valence	37/60
. . . increased/decreased policy support	44/28
Δ ex-ante utility	-0.00012
. . . as percent of difference between simultaneous elections and SP solution	-1%

Table 10: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.2$, $\sigma_v = 1$

	Sequential elections relative to simultaneous elections
Change of nominee	490
victory of an $a = 1$-candidate (change)	564(+316)
. . . increased/decreased valence	233/257
. . . increased/decreased policy support	195/196
Δ ex-ante utility	-0.004
. . . as percent of difference between simultaneous elections and SP solution	-4%

Table 11: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.4$, $\sigma_v = 1/3$

	Sequential elections relative to simultaneous elections
Change of nominee	277
victory of an a = 1-candidate (change)	614(+84)
... increased/decreased valence	88/189
... increased/decreased policy support	98/87
Δ ex-ante utility	-0.04
... as percent of difference between simultaneous elections and SP solution	-359%

Table 12: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.4$, $\sigma_\upsilon = 2/3$

	Sequential elections relative to simultaneous elections
Change of nominee	170
victory of an a = 1-candidate (change)	612(+37)
... increased/decreased valence	47/123
... increased/decreased policy support	64/50
Δ ex-ante utility	-0.028
... as percent of difference between simultaneous elections and SP solution	-306%

Table 13: 1000 simulated campaigns for $\sigma_\varepsilon = 0.5$, $\sigma_\eta = 0.4$, $\sigma_\upsilon = 1$

	Sequential elections relative to simultaneous elections
Change of nominee	249
victory of an a = 1-candidate (change)	648(+163)
... increased/decreased valence	132/117
... increased/decreased policy support	123/86
Δ ex-ante utility	+0.01
... as percent of difference between simultaneous elections and SP solution	+12%

Table 14: 1000 simulated campaigns for 3 candidates (one with $a = 0$ and two with $a = 1$, for $\sigma_\varepsilon = 0.25$, $\sigma_\eta = 0.2$, $\sigma_\upsilon = 2/3$