Primes, Strategic Voting, and Candidate Polarization

Jonathan Woon*
University of Pittsburgh

May 20, 2016†

Abstract

Do primary elections cause candidates to take extreme, polarized positions, impeding convergence to the median voter? Focusing on the strategic incentives arising from the interaction between electoral institutions and voting behavior, I address this question by directly comparing candidate positions in elections with and without primaries, taking advantage of the ability to manipulate and control electoral institutions theoretically and experimentally. Standard equilibrium analysis predicts full convergence to the median voter’s position, but behavioral game theory predicts divergence when players have out-of-equilibrium beliefs. The effect of primary elections depends on the strategic sophistication of candidates and voters, causing polarization when primary voters “sincerely” select extreme candidates. I find that candidate positions in the experiment diverge substantially from the median voter and that primaries cause greater candidate polarization, but only under some conditions. Instead of voting myopically, primary voters employ a strategy that weeds out candidates who are either too moderate or too extreme, generating ideological “purity,” thereby reinforcing polarization. The analysis highlights the importance of behavioral assumptions rather than preference divergence in understanding the effects of electoral institutions.

JEL Classifications: C72, D72

---

*Associate Professor, Department of Political Science, Department of Economics (secondary), and Pittsburgh Experimental Economics Laboratory, woon@pitt.edu, 4443 Wesley W. Posvar Hall, Pittsburgh, PA, 15260
†Thanks to Keith Dougherty, Sandy Gordon, Greg Huber, Scott Moser, Charlie Plott, Alan Wiseman, participants at the Yale CSAP American Politics Conference, seminar participants at Washington University in St. Louis, and members of the Pittsburgh Experimental Economics Lab for helpful comments and feedback. I am also indebted to Kris Kanthak for vigorous discussions during the early stages of this project. Previous versions of the paper were presented at the 2014 Annual Meeting of the American Political Science Association, the 2016 Public Choice Society Meeting, and the 2016 Midwest Political Science Association Conference. This research is supported by the National Science Foundation under Grant No. SES-1154739 and was approved by the University of Pittsburgh Institutional Review Board under protocol PRO14060001.
The partisan primary system, which favors more ideologically pure candidates, has contributed to the election of more extreme officeholders and increased political polarization. It has become a menace to governing.

— Sen. Charles Schumer (D-NY)

The divergence between candidates and legislators from the two major parties is an enduring feature of the American political landscape (Ansolabehere, Snyder and Stewart 2001, Bonica 2013, Poole and Rosenthal 1984, 1997), and the fact that polarization is at historically high levels is a significant concern for scholars and observers of democratic governance, representation, and public policy (Hacker and Pierson 2006, Mann and Ornstein 2013, McCarty, Poole and Rosenthal 2006, 2013). Indeed, politicians and the popular press often lay much of the blame on partisan primary elections. Blaming primaries is understandable, as there is a simple, intuitively appealing argument: Candidates take extreme positions because they must appeal to partisan primary voters, whose preferences are more extreme than those of voters in the general election.

There is some evidence to suggest that primary elections promote extremism, but overall the empirical record is mixed. Extremists are more likely to win congressional primaries than moderates (Brady, Han and Pope 2007), and legislators elected under closed primaries take more extreme positions than legislators elected under open primaries (Gerber and Morton 1998). But other analyses find that polarization is largely unrelated to the introduction of direct primaries (Hirano et al. 2010) and to the variation in the openness of primaries across states (McGhee et al. 2014). At best, primaries may cause polarization under limited circumstances (Bullock and Clinton 2011), and despite the divergence of candidate positions, general elections nevertheless exert nontrivial pressure on candidates to moderate (Hall 2015, Hirano et al. 2010).

How, then, should we understand the causal relationship between primary elections and candidate positioning? In this paper, I examine the causal connection, both theoretically and experimentally, by comparing elections with and without primaries while holding other

features of the electoral environment constant (e.g., preferences and information). The analysis focuses on a particular aspect of primary elections—how the introduction of voters in the candidate selection process affects strategic competition between parties—while abstracting away from many other considerations that might also affect polarization such as candidate valence, turnout, activists, or campaign contributions (Adams and Merrill 2008, Callander and Wilson 2007, Hirano, Snyder and Ting 2009, Hummel 2013, Meirowitz 2005, Snyder and Ting 2011).

My analysis follows a long tradition of using spatial voting models to understand elections, and although existing spatial models (Aronson and Ordeshook 1972, Coleman 1972, Owen and Grofman 2006) predict candidate divergence in elections with primaries (two-stage elections), they do so in isolation and do not compare them explicitly to elections without primaries (one-stage elections). These models also assume that general election outcomes are probabilistic, which is theoretically consequential because the mechanism producing divergence in these models is the combination of policy-motivations and uncertainty about which candidate will win the general election—the same forces that generate incentives for candidate divergence in the absence of primaries (Calvert 1985, Wittman 1983). Thus, it is unclear from these models whether polarization can be traced to any distinctive features of primaries per se, as electoral institutions. But by explicitly comparing institutions, my analysis generates insights regarding the connection between primaries and polarization, with an emphasis on the role of strategic expectations.

The setting for the investigation is extremely simple. I analyze a one-dimensional spatial voting model with candidates and voters who are both policy-motivated and have complete information about preferences. In contrast with existing models, my analysis shows explicitly that closed primaries can indeed affect the incentives of candidates to take different

---

2An exception is Jackson, Mathevet and Mattes (2007), who compare alternative nomination systems in a citizen-candidate framework. In their model, primary elections affect whose preference is decisive in nominating candidates and have no effect if party leaders and the median party voter have the same preferences. Other formal models of primary elections largely focus on considerations of voter uncertainty, incomplete information, and signaling along with issues of candidate valence and distributional concerns.
positions than they would in elections without primaries even when the preferences of a party’s candidates and voters are the same. The effect, however, depends critically on the expectations that candidates have about the strategic behavior of voters from the opposing party. Hence, my analysis identifies voter sophistication as a key explanatory variable.\(^3\) If a Democratic candidate expects Republican primary voters to behave sincerely and to vote for their party’s extremists, then the Democrat’s best response is to take a more extreme position than in the absence of primaries. On the other hand, if the Democrat expects Republican primary voters to prioritize winning the general election and to select moderates, then the Democrat’s response is to take more moderate positions as well.

Importantly, the conditional effect of primaries in my analysis is also only apparent when I relax the behavioral assumptions of standard game theoretic analysis. Under conditions of complete information, the model is uninteresting since Nash equilibrium predicts full convergence to the general election median voter in elections with and without primaries. This is because players have rational expectations in which beliefs and actions mutually reinforce one another (Aumann and Brandenburger 1995). In behavioral game theory, much of the theoretical apparatus from standard game theory is retained, including preferences, game forms, information, and (sometimes) optimization (Camerer 2003). One key difference is to allow players to have “incorrect” or “non-equilibrium” beliefs about others’ actions, rationality, or strategic thinking (Crawford, Costa-Gomes and Iriberri 2013). Players are nevertheless strategic in the sense that they best respond to what they think other players do, even though their beliefs may be incorrect (Camerer, Ho and Chong 2004, Nagel 1995, Stahl and Wilson 1995).\(^4\) By allowing voters and candidates to have incorrect beliefs, behavioral

---

\(^3\)Two-stage models, such as Owen and Grofman (2006), typically maintain the assumption that voters are strategic and forward-looking rather than sincere (myopic). Several formal models consider the issue of raiding and cross-over voting in open primaries (Cho and Kang 2014, Chen and Yang 2002, Oak 2006), which are outside the scope of my analysis. This result also differs from Adams and Merrill (2014) who find that strategic versus expressive voting both generate divergence, but in their model candidates are office-motivated and vary in their campaign skills.

\(^4\)Behavioral game theory need not be seen as incompatible with standard game theory. For example, standard equilibrium concepts can be viewed as generating predictions about behavior in the long-run, while behavioral game theory provides a framework for understand behavior in the short or medium-term before players learn or adapt.
game theory generates a rich set of competing predictions about the effect of primaries on candidate divergence.

In turning to the laboratory to test these predictions, I design and conduct two experiments in order to observe the choices that people actually make under controlled conditions that closely match the theoretical environment. Both experiments implement the game analyzed in the theoretical framework, but they differ with respect to ancillary design choices such as the assignment of player roles. Even though participants in the experiment are undergraduates, they are tasked with a straightforward problem of strategic reasoning in a spatial policy environment. They know their goal is to obtain an outcome closest to their ideal point, but that positions closest to the median voter are more likely to win the general election. This captures the key feature of the problem that real-world primary voters face, so the results provide insights about the way that real people confront this strategic trade-off.\footnote{See Woon (2012b) for a discussion of behavioral inference in incentivized experiments.}

Furthermore, laboratory control ensures that decision-makers face homogeneous conditions that minimize or eliminate the influence of various real-world confounds, such as differences in preferences, information, or the intensity of partisan attachments.

This paper brings together two distinct literatures in experimental political science. The experimental literature on candidate positioning in two-party elections finds a strong tendency for candidates and election outcomes to converge to the median voter’s position or, more generally, to the Condorcet winner under a variety of conditions, including incomplete information (Collier et al. 1987, McKelvey and Ordeshook 1982, McKelvey and Ordeshook 1985). The exception is when candidates are ideological and voting is probabilistic (Morton 1993). An equally robust finding from experiments on strategic voting, however, is that strategic or sophisticated voting is rare (Cherry and Kroll 2003, Eckel and Holt 1989, Herzberg and Wilson 1988, Plott and Levine 1978, Van der Straeten et al. 2010).\footnote{An exception is Smirnov (2009), who studies endogenous agendas and finds behavior consistent with sophisticated expected utility maximization.}
Taken together, previous studies suggest that voters will be myopic and that primary elections should therefore generate increased candidate polarization.

The experiment yields several key findings. I find that primary elections do cause greater polarization, but the effect is sensitive to the ancillary features of the experiment. In the first experiment, primaries have a polarizing effect only when there is no feedback available to subjects, while primaries have a consistent polarizing effect in the second experiment (with feedback) when candidate choices have greater salience. In all conditions of the experiment, candidates take positions that diverge significantly from the median voter regardless of whether the elections have one or two stages: two-stage elections are not necessary for candidate divergence. This finding lends support for the utility of behavioral game theory over standard Nash equilibrium analysis in this setting and is surprising in light of the existing experimental research on candidate positioning. With respect to voting behavior, I find that primary voters display a fair degree of strategic sophistication. They support neither party extremists nor party moderates unconditionally. Instead, they select candidates with intermediate positions—consistent with their own subjective beliefs about optimal candidate positions, which tend to be approximately halfway between the median voter and their own party’s ideal point. This behavior generates a greater concentration of candidate positions around an average position that diverges from the median voter, and such ideological “purity” has the effect of bolstering candidate polarization.

**Theoretical Framework and Analysis**

I consider an environment with two parties, Party $L$ and Party $R$, competing to win a single office. Candidates choose positions in a one-dimensional policy space, and the winning candidate’s position is implemented as the policy outcome. In the electorate, there are an equal number of voters in each party and a set of independent, non-partisan “swing” voters. Candidates and voters alike are entirely *policy-motivated*, caring only about the
location of the policy outcome \( w \in \mathbb{R} \). The incentive to win office is therefore purely instrumental in this model, which departs from usual Downsian office motivations. Parties are completely homogeneous in that candidates and voters belonging to the same party are identical and have the same ideal point. Thus, there are effectively three ideal points: \( \theta_L \) for members of Party \( L \), \( \theta_R \) for members of Party \( R \), and \( \theta_M \) for the electorate’s median voter, where \( \theta_L < \theta_M < \theta_R \). I assume that preferences are symmetric and single-peaked. Specifically, in the experimental implementation, all actors have linear loss utility functions, 
\[
u_i(w) = K - |w - \theta_i|, \quad \text{for} \quad i \in \{L, M, R\} \quad \text{and some constant} \quad K.
\]
Preferences are also common knowledge, so the election takes place under conditions of complete information.

There are two types of elections. In one-stage elections (1S), there is one candidate from each party whose positions are \( c_L \) and \( c_R \), respectively, and one round of majority rule voting to select the winning candidate. In two-stage elections (2S), there are two candidates from each party (denoted \( c_{L1} \) and \( c_{L2} \) from Party \( L \), \( c_{R1} \) and \( c_{R2} \) from Party \( R \)) who compete in an intra-party first round election (the “primary” election). The two candidates who win their respective party primaries then compete in a second round election (the “general” election) to select the winning policy \( w \). In other words, the parties hold simultaneous “closed” primaries so that the voter with ideal point \( \theta_L \) effectively chooses \( c_L \in \{c_{L1}, c_{L2}\} \) in the Party \( L \) primary at the same time that the voter with ideal point \( \theta_R \) chooses \( c_R \in \{c_{R1}, c_{R2}\} \) in the Party \( R \) primary. In the general election, the median voter with ideal point \( \theta_M \) chooses the election outcome from the two candidates selected by the parties’ respective median voters, \( w \in \{c_L, c_R\} \).

To generate predictions about candidate positioning and to identify the potential effects of the election format, I consider a variety of alternative behavioral assumptions. I begin with standard game theoretic analysis, applying Nash equilibrium as the standard solution concept. Because I am interested in making behavioral predictions, the interpretation of Nash equilibrium is worth a brief discussion. One way to interpret Nash equilibrium is to think of it as embodying an idealized set of assumptions such that actors are not only fully rational
but also that their rationality is common knowledge (Aumann and Brandenburger 1995). In this interpretation, we can think of political actors as forming beliefs about others’ current and future behavior that are fully consistent with players’ actual strategies and behavior. Alternatively, Nash equilibrium can be interpreted as merely representing a stable outcome in which strategies are mutual best responses without necessarily invoking an epistemic or belief-based justification of how individuals make decisions in games. Such an approach, however, does not make clear cut predictions about how games are played before an equilibrium state is reached. Nevertheless, under a wide variety of learning models, experience can lead play to converge to Nash equilibrium (Fudenberg and Levine 1998), and the role of experience can be investigated experimentally.

Relaxing the Nash assumption of the mutual consistency of beliefs and actions generates an interesting variety of behavioral possibilities. In my analysis, I first explore the implications of voter sophistication for candidate positioning while holding candidate rationality constant. If voting is “sincere,” then primary elections produce more polarized candidates than voting that follows an equilibrium strategy. I then consider another departure from standard assumptions: beliefs that some players make mistakes in choosing their positions. They might do so for any number of reasons, such as miscalculating the optimal position, misjudging or underestimating the rationality of others, or having preferences over outcomes of the game that are not fully captured by their material payoffs. Strategically sophisticated players, recognizing that there are other players who make mistakes, will then choose positions that differ from the Nash predictions—in the direction of their parties’ ideal points—but that are optimal given their own beliefs about the distribution of opponents’ positions. Introducing noise or the possibility of mistakes generates divergence in both one-stage and two-stage elections.

With noise, the effect of introducing a primary election is more complicated. Similar to the case in which candidates do not make mistakes, the optimal positions depend critically on voting behavior. If voters always choose moderate primary candidates, then two-stage
elections will generate greater convergence of candidate positions than in one-stage elections. However, if voters always choose more extreme primary candidates, then candidates in two-stage elections will be more polarized than candidates in one-stage elections. There is also a third possibility. If voters form their own beliefs about the position most likely to maximize their expected utility and vote for candidates closest to this position, then the degree of candidate divergence in two-stage elections is increasing in what we might call voters’ belief-induced ideal points. Furthermore, there exists a belief-induced ideal point such that candidates’ optimal positions diverge from the median voter by the same amount in both one-stage and two-stage elections. Behavioral game theory establishes a critical link between the effect of primaries and candidates’ beliefs about opponents’ primary voting behavior.

Candidate equilibrium with fully strategic voters

Subgame perfection makes identical predictions for both one-stage and two-stage elections: In any equilibrium, the winning candidate’s position is the median voter’s ideal point. In one-stage elections, the logic is straightforward. The median voter chooses the party candidate closest to his or her ideal point as the winning candidate, so if one candidate adopts $\theta_M$ as a campaign position, no other position can defeat it. In the unique equilibrium of the one-stage election game, both parties’ candidates must choose $c_L = c_R = \theta_M$. If not, either the winning party’s candidate could do better by finding a position closer to her ideal point while still winning the election or the losing candidate can find a position that wins the election, thereby obtaining a better policy outcome for herself. Thus, $w = \theta_M$ is the unique equilibrium policy outcome.

In two-stage elections, the outcome is the same, but the equilibrium strategies of the primary voters must be specified. Given a set of candidate positions and voters’ expectations that the general election median voter will choose the more moderate of the parties’ candidates, a primary voter’s strategy is to choose the candidate closest to her ideal point as long as she believes the candidate will also win the general election (and in equilibrium,
the voter’s beliefs about which candidate will win are correct). Because candidates and voters have the same preferences, the incentives guiding optimal candidate strategies in the one-stage election are the same as those that guide rational voting behavior in two-stage elections: if offered the same choices, candidates and voters would choose the same position (the only difference is that candidates can choose any position while primary voters’ choices are constrained).

In any equilibrium of the two-stage election game, there must be at least one candidate from each party located at \( \theta_M \), so primary voters will always be observed choosing the moderate candidate along the path of play. If so, both parties’ primary voters will select a candidate at the median voter’s ideal point and the policy outcome is therefore \( w = \theta_M \). Ruling out other possible outcomes then follows from the same logic as in the nonprimary election. Fully strategic behavior from voters predicts full convergence to the median voter’s position in both one-stage and two-stage elections.

**Prediction 1.** If voters and candidates are rational, forward-looking agents and form correct beliefs about others’ behavior, then (a) moderate candidates from each party will adopt the median voter’s position and (b) primaries will have no effect on the polarization of candidates in the general election.

**Candidate equilibrium with sincere voters**

I next consider the possibility that primary voters are myopic and vote “sincerely.”\(^7\) I assume that sincere voters simply vote for the candidate closest to their ideal points, so they are myopic in the sense that they fail to recognize that the candidate’s chances of winning the general election affect the policy outcome (and hence their payoffs). With myopic voters, the two-stage election game has multiple equilibria in which candidates take divergent positions

\(^7\)While the overall level of voter rationality remains an ongoing subject of academic debate, the assumption that voters are myopic is consistent with recent observational and experimental research (e.g., Healy and Malhotra 2009, Huber, Hill and Lenz 2012, Woon 2012a). A theory of elections with boundedly rational, behavioral voters is also worked out by Bendor et al. (2011).
while the equilibrium of the one-stage election game remains the same (full convergence, since there are no primary voters).

In any equilibrium of the two-stage election game with sincere voters, candidates within each party must adopt the same position, and opposing party candidates must be equidistant from the median voter. Specifically, an equilibrium is characterized by the condition that $c_{L1} = c_{L2} = \theta_M - \delta$ and $c_{R1} = c_{R2} = \theta_M + \delta$, where $\delta > 0$ denotes some amount of divergence between candidates. The median voter’s strategy is to select the candidate closest to her own ideal point, breaking ties in favor of each party with equal probability.\(^8\) The result of the general election is therefore a lottery over $w \in \{\theta_M - \delta, \theta_M + \delta\}$, and the expected value of the outcome is the median voter’s position, $E[x] = \theta_M$. Any candidate who adopts a more extreme position would, at best, be able to win their own primary but then would lose the general election with certainty. Moving to a more moderate position would not change the result of the primary and thus would not change the general election result either. Since no candidate can obtain a better policy outcome by unilaterally adopting a different position, campaign promises characterized by intra-party convergence and inter-party symmetric divergence constitute an equilibrium of the primary election game with sincere voters. The basic intuition underlying this result is that because of sincere primary voters, intra-party competition limits any one candidate’s ability to moderate their party’s position in the general election. Thus, in contrast to full convergence in one-stage elections, any amount of divergence can be supported in two-stage elections.

**Prediction 2.** If candidates are rational and forward-looking but primary voters “sincerely” select candidates closest to their own ideal points, then (a) candidates from each party will take positions that diverge from the median voter by the same amount in two-stage elections, and (b) winning candidates will be at least as polarized in two-stage elections as in one-stage elections.

\(^8\)Note that it is also possible to construct equilibria in which the median voter has a bias for one of the parties (i.e., breaks ties in favor of one party rather than randomizing), but this would not affect the equilibrium positions of the candidates. Thus, even though the random tie-breaking rule matches the experimental setup, it is not necessary for the results.
Candidate best responses to out-of-equilibrium beliefs

The previous sections assumed that candidates correctly anticipate whether voters use Nash or sincere voting strategies and that their beliefs about other candidates are consistent with those candidates’ actual behavior. The mutual consistency of candidates’ beliefs and actions might break down in a number of ways. Candidates are likely to face cognitive constraints, they may engage in incomplete strategic reasoning, or they may doubt the rationality of other candidates. In this section, I apply the notion of limited strategic sophistication motivated by level-k models in behavioral game theory (Crawford 2003, Nagel 1995, Stahl and Wilson 1995), positing that candidates have some (possibly arbitrary) beliefs and analyze the best response to such beliefs.

To model this, let candidate $c_i$’s beliefs about the positions of candidates from the opposing party $j \neq i$ be given by the cumulative distribution $F(c_j)$, with full support bounded by the median voter $\theta_M$ and the ideal point of the opposing party $\theta_j$. The distribution $F(c_j)$ can be interpreted as an objective probability distribution if candidates’ choices are noisy and $F(c_j)$ reflects the true distribution of candidate positions. Alternatively, $F(c_j)$ can be interpreted as $i$’s subjective beliefs about the set of opposing candidates. Regardless of the interpretation, as long as candidate $c_i$ believes that there is some chance that $c_j$ is between $\theta_M$ and $\theta_j$, it is straightforward to show that a policy-motivated candidate will choose a position that diverges from the median voter’s ideal point.

In the case of a candidate from party $R$, the optimal position maximizes

$$EU(c_R) = \int_{\theta_L}^{2\theta_M-c_R} u(c_R) f(c_L) dc_L + \int_{2\theta_M-c_R}^{\theta_M} u(c_L) f(c_L) dc_L$$

(1)

where the integral on the left is the expected utility if $c_R$ is closer to the median voter and wins while the integral on the right is the expected utility if the opposing candidate $c_L$ is closer to the median. Note that because the median voter will always choose the moderate candidate, the choice of $c_R$ is governed by three considerations. First, and more obviously, $c_R$
is the policy outcome if $R$ wins ($w = c_R$, given by the first integral). Second, it sets a limit to how far the winning policy can be from $R$’s ideal point $\theta_R$ if $R$ loses ($|w| = |c_L| < |c_R|$, represented by the second integral). Third, it affects the probability of winning (the upper limit of the first integral and the lower limit of the second). To illustrate this concretely with a specific functional form, suppose that $\theta_R = 1$, the left party’s ideal point is $\theta_L = -1$, the median is $\theta_M = 0$, and $F(c_L)$ is a uniform random variable, $c_L \sim U[-1, 0]$. With linear loss utility, the optimal position that balances these considerations is $c^*_R = \frac{1}{3}$. This is illustrated by the solid line in Figure 1 showing $EU(c_R)$.

How do beliefs about positions interact with the election format and expectations about primary voting behavior? The effect of primaries will depend on the candidates’ beliefs about the opposing party’s primary voters. In a one-stage election, a candidate expects to face a single opposing candidate $c_j$ randomly drawn from $F(c_j)$, and the optimal position against a single draw is $c^*_R$ as in the example above. In a two-stage election, however, two candidates, $c_{j1}$ and $c_{j2}$, are drawn from the distribution $F(c_j)$, but which of these candidates moves on to the general election depends on party $j$’s primary voters. Denote the distribution of the general election opponent by $G(c_j)$, which will typically be different from the distribution of the opposing party’s primary candidates $F(c_j)$.
More precisely, suppose that both of the opposing party’s candidates are independently drawn from $F(c_j)$. If $j$’s primary voters unconditionally select the more moderate candidate (as they would in equilibrium), then party $j$’s general election candidate will be the more moderate of two independent draws from $F(c_j)$. Applied to party L, $c_L = \max\{c_{L1}, c_{L2}\}$, from which it follows that the distribution of $c_L$ in the general election is $G^m(c_L) = \Pr(\max\{c_{L1}, c_{L2}\} < c_L)$. This distribution first order stochastically dominates $F(c_L)$, which implies that the optimal position $c^m_R$ will be more moderate than $c^*_R$. Again consider the specific case where $\theta_M = 0$, $\theta_R = 1$, and $c_L \sim U[-1,0]$. Since $c_{L1}$ and $c_{L2}$ are independent uniform random variables, the distribution of general election candidates has the PDF $g^m(c_L) = 2c_L + 2$ for $c_L \in [-1,0]$. Substituting $g^m(c_L)$ for $f(c_L)$ in (1) yields the optimal position $c^m_R = \frac{1}{5} < c^*_R = \frac{1}{3}$. This is illustrated by the lower dashed line showing $EU^m(c_R)$ in Figure 1. The expectation that the opposing party’s primary voters prefer moderate candidates therefore generates a greater incentive for moderation in two-stage elections than in one-stage elections.

Alternatively, if party $j$’s primary voters unconditionally select the more extreme candidate (as they would if they voted sincerely), then primaries generate an incentive for extremism in two-stage elections. This is because the primary election selects opposing candidates who are more likely to be extreme than a single random draw from $F(c_L)$. In this case, $c_L = \min\{c_{L1}, c_{L2}\}$, and the distribution is $G^e(c_L) = \Pr(\min\{c_{L1}, c_{L2}\} < c_L)$, which is first order stochastically dominated by $F(c_L)$. Hence, $c^e_R > c^*_R$. Maintaining the previous functional form assumptions, the PDF of this distribution is $g^e(c_L) = -2c_L$ for $c_L \in [-1,0]$, and solving (1) given this distribution yields the position $c^e_R = \frac{1}{\sqrt{5}} > c^*_R$. The upper dashed line in Figure 1 shows $EU^e(c_R)$. Thus, the optimal position is more extreme than in the one-stage election.

Notice that if primary voters have their own beliefs about the distribution of the opposing party’s candidates, an unconditional voting strategy (either for the more extreme or the more moderate candidate) is myopic. A sophisticated voter does better by using a
different strategy. Suppose that, like candidates, $i$’s primary voters have beliefs given by $G(c_j)$. If so, then the voter’s induced preferences over primary candidates are also given by expected utility function in (1). Faced with the choice between 0 and $c_R^*$, the voter prefers the more extreme candidate. However, if the choice is between $c_R^*$ and 1, the voter prefers the more moderate candidate. Thus, there is a third characterization of voting behavior to consider: *sophisticated* primary voters favor the candidate that maximizes the expected utility function in (1) given their beliefs $G(c_j)$. Sophisticated primary voters recognize there is uncertainty in position of the opposing party’s general election candidate and behave accordingly.

To illustrate the effect that sophisticated primary voters have on candidate positioning, suppose that $L$’s primary voters have beliefs such that $c_L^* = -\frac{1}{2}$ is their *belief-induced ideal point* (so that they always vote for the primary candidate closest to $c_L^*$).\(^9\) If $c_{L1}$ and $c_{L2}$ are independently drawn from $U[-1,0]$, then the distribution of $L$’s general election candidate is described by the PDF $g^s(c_L) = 2 - 4|c_L + \frac{1}{2}|$. This distribution is triangular with the highest density at the induced ideal point $c_L^* = -\frac{1}{2}$. Using the same parametric assumptions as above, $R$’s optimal position is $c_R^* = \frac{1}{\sqrt{10}}$. This position is only slightly more moderate than the optimal position in one-stage elections, which suggests that there may exist (via an intermediate value theorem argument) an appropriate primary voting rule or distribution of beliefs $G(c_L)$ such that the optimal candidate position is equivalent in one-stage and two-stage elections.

The analysis thus demonstrates that when candidates take heterogeneous positions, the effect of primary elections on candidate positioning depends critically on the behavior of the *opposing* party’s primary voters. This effect operates through the selection of opposing candidates for the general election.\(^\text{10}\) Summarizing the theoretical analysis, I show that

---

\(^9\)This behavioral voting rule is a simplification. It does not strictly follow from the previous analysis since $EU(c_i)$ is not necessarily symmetric and $c_L^* \neq -\frac{1}{2}$ in the case of uniform beliefs. Nevertheless, it captures the main idea that primary voters prefer candidates with positions in the interior of the candidate space. Moreover, there must exist some distribution of beliefs for $F(c_R)$ such that $c_R^*= -\frac{1}{2}$.

\(^\text{10}\)Note that in a primary election, intra-party competition does not affect $c_i$’s best response. The reason is that candidates are policy-motivated and the position that maximizes (1) is weakly better than any other
given a distribution of primary candidates $F(c_j)$, the behavior of primary voters is crucial for determining the distribution of general election candidates $G(c_j)$. The distribution $G(c_j)$, in turn, affects the optimal candidate position $c_i^*$. As the distribution $G(c_j)$ changes such that extreme candidates become more likely, then the best response $c_i^*$ becomes more extreme.

Note that this analysis depends on the assumption that the distribution of candidate positions, $F(c_j)$, is the same in both one-stage and two-stage elections. This can be justified with a level-$k$ argument by modeling variation in cognitive levels or degrees of strategic reasoning. In a level-$k$ model, the least sophisticated candidates are designated level-0 and behave naively, and all other types best respond to the types that are one level below them. That is, level-1 types best respond to level-0, and in general level-$k$ players best respond to level-$k$-1 players. In my setting, it is consistent with a level-$k$ model to posit that level-0 candidates choose positions randomly, following the distribution $F(c_j)$. Level-1 candidates are somewhat more sophisticated, as they believe other candidates’ positions follow the level-0 distribution (irrespective of the election format) and best respond to them. Since the optimal solution to equation (1) yields the best response of a level-1 candidate, my analysis can therefore be interpreted as focusing on level-1 candidates. This provides a justification for assuming that the distribution of positions $F(c_j)$ is the same in one-stage elections as it is in two-stage elections, since level-1 beliefs do not depend on the type of election. Taking the level-$k$ analysis one step further, however, the positions of level-2 types will vary across institutions with differences dictated by the type of voting behavior as presented in the foregoing analysis.

In contrast to standard equilibrium analysis, which predicts full convergence, a simple model with non-equilibrium beliefs generates divergence in candidate positions, even in the position. To see why, suppose that the other candidate from party $R$’s location is $c'_R$. If $c'_R < c^m_R$ and $R$’s voters select moderates, then $c'_R$ will win the primary. Moving to any position $c''_R < c'_R$ (including $c^m_R$) yields lower expected utility while any position $c''_R > c'_R$ (including $c^m_R$) yields the same expected utility and hence is a best response. If the other candidate takes a more extreme position, $c'_R > c^m_R$, then $c^m_R$ wins the primary and maximizes expected utility. A similar argument applies to sincere primary voters. No matter what position the other candidate from one’s own party takes, adopting a position other than the one that maximizes (1) yields a worse expected payoff.
absence of primaries and with complete information about preferences. Furthermore, I have shown how the effect of primaries varies with candidates’ expectations about the opposing party’s voting behavior. Primaries can indeed cause greater polarization, but only if primary voters select sufficiently extreme candidates. They can also cause greater moderation. My analysis also suggests that the effect of primaries depends more on beliefs about opposing candidates’ positions than they do on the extremity of voter preferences, which might help explain why the empirical literature seems to find little or no effect of the openness of primaries on candidate and legislative polarization.

**Prediction 3.** If candidates have non-equilibrium beliefs about the distribution of opposing candidates, then (a) candidate positions will diverge from the median voter’s ideal point in both one-stage and two-stage elections, (b) the direction of the effect of primary elections on candidate polarization depends on expectations about voting behavior, and (c) polarization in two-stage elections is increasing in the expected extremity of candidates selected by the opposing party’s primary voters.

**Experiment 1**

**Procedures**

Experiment 1 was designed primarily to test the predictions about candidate positions and to maximize the number of observed campaign promises made by each subject.\(^{11}\) Throughout the experiment, the policy space comprises the set of integers from 1 to 200, and payoffs are given by the linear loss function \(200 - |w - \theta_i|\), where \(w\) is the winning policy position and \(\theta_i\) is player \(i\)’s ideal point.\(^{12}\) Payoffs are denominated in “points” and converted to cash by

---

\(^{11}\)See the Appendix for the full text of the instructions for the 2S treatment.

\(^{12}\)I chose linear loss for two reasons. One is that it is simple to describe and easier to explain to subjects than a quadratic loss function. The other is that with a linear loss function, every possible policy outcome between the parties’ ideal points generates an equal amount of total social welfare. In contrast, quadratic utility implies that social welfare is maximized at the midpoint between the parties. The use of linear loss utility therefore makes it unlikely that subjects will try to maximize total social welfare as their objective.
dividing by 10 and rounding to the nearest quarter; each election is worth between $0 and $20 dollars. The final payment is determined by randomly selecting one election to count for payment (thereby minimizing wealth effects) and adding a $7 show-up fee.

Subjects participate in multiple elections, and the instructions emphasize that each election is to be treated as a “separate decision task.” Each round of voting in the 2S election is referred to as a separate “voting stage” rather than as an election. For each election, subjects are divided into two groups, and every member of the group has the same ideal point (referred to as “favorite positions” in the experiment). I chose identical within-group ideal points in the experimental design to rule out preference heterogeneity as a cause of polarization in the 2S election treatment, focusing instead on the potential for strategic voting. The numerical values of the ideal points varied from election to election, while the exact sequence is identical across sessions and treatments. I varied the numerical values in order to encourage subjects to pay attention and think about their relative, rather than absolute, positions. For the most part, however, I also keep the distance between parties fixed from election to election to maintain the homogeneity in the strength of incentives throughout the experiment, varying this distance only a few times to test whether subjects respond to a change in the level of preference polarization. Specifically, the groups’ ideal points are located symmetrically about the median voter’s position, and in half of the elections they are located \( \pm 50 \) positions from the median (“baseline” polarization) and in the other half are \( \pm 75 \) positions from the median (“increased” polarization).

In addition to the human players, there is one “computer voter” who has a distinct ideal point and, as the instructions explain to subjects, is “like a robot programmed to

---

13 We can think of each group as a party, although I am careful to avoid using the term “party” when describing the game to subjects. For 2 sessions of each treatment, groups were randomly assigned before every period (random matching), and in the other sessions groups were assigned before the first election and remained fixed in every subsequent election (fixed matching). There do not appear to be appreciable differences between the results for random and fixed matching, so for the analysis I ignore this distinction and pool the data by type of election.

14 To determine the sequence of values, I randomly selected the median’s position, \( \theta_m \), from the integers between 51 and 150 for the moderate polarization elections and between 76 and 125 for the high polarization elections.
always vote for the candidate whose campaign promise gives it the higher payoff value.”\textsuperscript{15} We can think of the computer voter as non-partisan, as it only participates in the 1S election and in the second stage of the 2S election. In the case of ties, the computer votes for each candidate with equal probability. The computer voter is the median voter in every election, and because there are equal numbers of human voters in each group, the computer voter is decisive if human voting falls along strict party lines. During each election, subjects know the computer voter’s ideal point, as well as their own ideal point and the ideal point of the other group, thereby satisfying the complete information assumption.

The procedures for each election are as follows. At the beginning of the election period, subjects first learn the position of every player’s ideal point. Every subject then chooses a “campaign promise” (their policy position). They know that if their campaign promise is selected as the winning position, it affects every player’s payoff. After subjects choose their campaign promise, the computer randomly selects candidates from each group: one candidate from each group in the 1S election and two candidates from each group in the 2S election, with each group member equally like to be selected and the selection of candidates independent across election periods. The rest of the subjects are assigned to the role of a voter in that election. Thus, at the beginning of each election, every subject is a potential candidate and does not know whether he or she is a candidate until after submitting a campaign promise.

This method of role assignment is similar in spirit to the strategy method and maximizes the number of observed positions in the experiment given that one of the primary goals of the experiment is to measure and test candidate positioning behavior. At the same time, I needed to have a large enough number of voters in each group to create uncertainty about the behavior of primary voters. One concern with this feature of the design is that some subjects might mistakenly believe that their choice is inconsequential because there is either a 1/7 or 2/7 chance that they will be a candidate. While sophisticated subjects would

\textsuperscript{15}This is similar to language used by Morton (1993).
recognize that the probability of being selected has no effect on the structure of incentives (as it only scales the magnitude of the expected payoff) and hence should have no effect on their choice, I took some precaution to ensure subjects’ understanding would be more uniform by explicitly instructing them that “you should choose your campaign message as if you were actually selected as a candidate.” The design of Experiment 2 addresses these concerns by assigning subjects to fixed roles throughout the experiment.

Once the candidates are selected, the game proceeds to the voting stages. In the 2S election, voters first choose between one of their group’s two candidates by majority rule. Each primary (first stage) vote is held simultaneously, and neither party knows the positions of the other group’s candidates while voting. Abstentions are not allowed (and since there are five voters, there is no need for a tie-breaking rule). After each group selects its nominee, a second round of voting takes place to choose the winning policy from the two groups’ nominees. All voters participate in this second round—effectively the “general election”—along with the computer voter. In contrast to the 2S election treatment, the 1S election treatment features only one round of voting in which every voter (including the computer voter) participates.

In each experimental session, there are a total of 40 elections divided into two parts, and each part varies the type of feedback subjects receive. There are 10 elections in Part 1, and subjects receive no feedback about the results of any of the elections in Part 1. This generates multiple observations of behavior under conditions in which subjects have to reason introspectively, forming their own strategies and beliefs without being able to learn

---

16One might be concerned that such instructions might be “heavy handed,” but note that the instructions do not tell subjects what positions to choose, only that they should think of their choice as being equally consequential. Such instructions therefore enhance control by increasing the uniformity of subjects’ understanding of the game, thereby allowing the analysis to focus on interesting variation in terms of subjects’ strategic expectations. Notice also that because voters and candidates have identical preferences, this aspect of the design should cause no problems for subjects interpreting the game differently as voters than as candidates.

17To avoid priming subjects’ political attitudes regarding primaries, I avoid referring to the two rounds of voting as a “primary” and “general” election but instead refer to them as the “first voting stage” and the “second voting stage.”
Table 1: Structure of Experiment 1

<table>
<thead>
<tr>
<th>Elections</th>
<th>Condition</th>
<th>One-stage</th>
<th>Two-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>No feedback, moderate polarization</td>
<td>420</td>
<td>490</td>
</tr>
<tr>
<td>6-10</td>
<td>No feedback, high polarization</td>
<td>420</td>
<td>490</td>
</tr>
<tr>
<td>11-25</td>
<td>Feedback, moderate polarization</td>
<td>1260</td>
<td>1470</td>
</tr>
<tr>
<td>26-40</td>
<td>Feedback, high polarization</td>
<td>1260</td>
<td>1470</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Sessions</th>
<th># Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>98</td>
</tr>
</tbody>
</table>

about the behavior of others from past play.\textsuperscript{18} Within Part 1, elections 1-5 use ideal points with the baseline level of preference divergence (± 50 positions), and elections 6-10 use ideal points with increased polarization (± 75 positions). In Part 2, there are 30 elections, and subjects receive feedback about the election immediately after each election concludes (and before beginning the next election). The information subjects receive includes the positions of the subjects who were selected as candidates, the number of votes for each candidate, the winning position, and the payoff from the final outcome.\textsuperscript{19} As in Part 1, the first half of the elections in Part 2 (11-25) feature the baseline level of preference divergence while the second half (26-40) feature increased polarization.

Table 1 summarizes the structure of Experiment 1. The between-subjects manipulation varies the electoral institution to test the causal effect of primary elections given

\textsuperscript{18}The fact that the game is sequential and voters observe the positions of candidates when they vote presents some difficulty in completely preventing feedback between elections. That is, it would be impossible to prevent learning across elections if subjects completed each election game before proceeding to the next. I solved this problem by implementing a procedure similar in spirit to the strategy method. Instead of playing each election game one after another, subjects first chose their campaign promises for all 8 elections, then voted in all primaries for which they were assigned to be a voter, and then finally voted in all general elections in which they were a voter. In this way, subjects choose campaign promises without observing any of the positions of the opposing party or whether primary voters are strategic.

\textsuperscript{19}Subjects did not observe the full distribution of positions, only the positions of subjects selected as candidates.
policy-motivated candidates. The two within-subjects manipulations vary feedback and polarization. Comparing behavior in the no feedback condition (elections 1-10) against behavior in the feedback condition (elections 11-40) provides a test of whether polarization arises from introspective beliefs and strategic understanding of the game players (without feedback) or from experience and learning over time (with feedback, e.g., that candidates positions are noisy or that voters are myopic). Varying the degree of preference polarization tests whether polarization in candidate positions is responsive to the political actors’ underlying preferences.

The computer interface was programmed and implemented using z-tree (Fischbacher 2007), and Experiment 1 was conducted at the Pittsburgh Experimental Economics Laboratory (PEEL) between November 2014 and April 2015. There were 13 sessions: 6 of the 1S election treatment and 7 of the 2S election treatment, with 14 participants in each session (for a total of 182 participants). Each session took about about an hour and a half to complete, and subjects earned an average of $21.10 (including their show-up fee).

Electoral Dynamics

To get a sense for the kinds of promises candidates make and whether moderates or extremists win elections in Experiment 1, Figure 2 presents the sequence of candidate positions and outcomes for several sessions (2 one-stage sessions and 2 two-stage sessions). The horizontal axis shows each election, and the vertical axis shows the promises of the subjects selected as the candidates. These positions are median-adjusted so that the general election median voter’s position is 0. General election candidates are depicted using solid shapes (candidates in one-stage elections and the primary winners in two-stage elections) while primary candidates who lost the first stage election are depicted with hollow shapes. The winning position of the general election is shown by the solid line. Although the dynamics of each session differ, these plots reveal several noteworthy patterns.
Figure 2: Sample session dynamics
First, the positions of candidates from the two parties clearly diverge from the median voter’s position. This is true for both one-stage and two-stage elections, and it appears to persist over the course of the experiment even after subjects gain considerable experience. In session 10 (one stage), for example, the candidates from each party choose positions close to their own ideal points, and the polarization between the candidates’ positions increases when the underlying preference polarization increases. Along with divergence, there also appears to be substantial heterogeneity and fluctuation in candidate positions.\footnote{The figures also reveal that candidates and voters sometimes make mistakes. For example, in election 1 in session 4, both parties’ candidates are located to the left of the median voter, with the party R candidate located at leftmost position in the policy space.}

Second, while the general election candidate closer to the median voter’s position generally wins, it is rare for the winning candidate to be located exactly at the predicted equilibrium position. Even in session 4 (one stage), in which the electoral outcome appears most frequently near the median voter’s position, the winning candidate is located at the median’s position in only 3 elections (in another 8 elections, the winning candidate is \(\pm 1\) from the median voter’s position). In session 10, the winning candidate usually appears to be just barely closer to the median voter than the losing candidate.

Third, primary voters sometimes select more extreme candidates and sometimes select more moderate candidates. Notably, there are several candidates in two-stage elections who locate at exactly the median voter’s position yet lose the primary. In session 1, there were 11 out of 14 such candidates, and in session 6, there were 6 out of 10. While this could suggest that primary voters prefer extremists, there are also many elections in which the more moderate candidate wins. For example, in election 11 of session 6, the left party candidate at -30 defeated the candidate at -45, and the right party candidate at 20 defeated the candidate located at 50, with the right party candidate (who is closer to the median voter) winning the general election. Indeed, Figure 2 depicts losing candidates in primary elections on either side of the parties’ winning candidates (indicated by the fact that the hollow candidate markers appear both above and below the solid ones).
These sample dynamics suggest that standard game theoretic analysis poorly predicts candidate positions and voting behavior in the experiment. Whereas equilibrium predicts complete candidate candidate convergence in both one-stage and two-stage elections, I find that candidates’ positions diverge. The considerable heterogeneity in candidate positions and the selection of extreme candidates by primary voters indicate that behavioral game theory and non-equilibrium analysis may be useful tools for understanding the consequences of electoral institutions. Of course, Figure 2 only provides a snapshot of experimental behavior. The remainder of the analysis demonstrates that many of the patterns described above generalize across subjects and sessions from Experiment 1.

Candidate Positions

Figure 3 shows the average positions over time and by election format for all candidates (top panel) and for winning candidates (bottom panel). In the remainder of the analysis, I measure the extremity of a candidate’s position (vertical axis) by normalizing positions so that a subject’s own ideal point is 1 and the median voter’s ideal point is 0 (the opposing party’s ideal point is $-1$ on this transformed scale). The top panel of Figure 3 shows that candidate positions clearly diverge from the median voter’s position throughout Experiment 1 regardless of the election format. This divergence also appears to persist over time and with no apparent effect of primary elections on polarization. The average normalized position across all rounds is 0.452 in the 1S condition and 0.456 in the 2S conditions. Subjects choose positions only slightly closer to the median voter than the midpoint between their group’s ideal point and the median voter’s ideal point. While the bottom panel shows less stability in the positions of winning candidates due to the fact that there are a small number of sessions per treatment, there are some differences conditional on the availability of feedback. Without feedback, there is some convergence of winning candidates to the median voter’s position in elections without primaries and an increase in divergence once feedback is introduced in
Figure 3: Average positions and outcomes in Experiment 1
Table 2: Regression analysis of positions in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>No feedback (elections 1-10)</th>
<th>Feedback (elections 11-40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Party</td>
</tr>
<tr>
<td>Primary (2S) Elections</td>
<td>0.004</td>
<td>0.175*</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>Increased Polarization</td>
<td>0.081</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Experience</td>
<td>-0.008</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.387**</td>
<td>0.400**</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>N</td>
<td>1820</td>
<td>260</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.001</td>
<td>0.028</td>
</tr>
</tbody>
</table>

* $p < .05$ ** $p < .01$, OLS regressions with robust standard errors in parentheses, clustered by subjects in (1), (4) and sessions in (2), (3), (5), and (6).

In elections with primaries, however, the positions of winning candidates remain polarized throughout the experiment.

Table 2 presents a series of ordinary least squares regressions to measure the effect of primaries on candidate divergence while controlling for feedback and experience. I estimate separate models for each type of position data produced by the experiment and by the availability of feedback. Columns (1) and (4) present models for the positions chosen by all subjects, columns (2) and (5) for the positions of the party candidates in the general election, and columns (3) and (6) for the winning candidates. Increased Polarization is a dummy variable indicating baseline (0) or increased (1) polarization, and the Experience variable is a count of the number of previous elections (starting at 0).

The results of the regression analysis generally reinforce the visual interpretation of the data displayed in Figure 3. Positions are divergent (as measured by the intercept) and do not change over time (as the coefficients on Experience are small and insignificant across the models). Although primary elections have no effect on the positions chosen by
all candidates, they do have a statistically significant effect on the divergence between party candidates (those standing in the second voting stage) in the absence of feedback. In 1S elections, the divergence of party candidates from the median voter is 0.4 on the normalized scale (i.e., 40% of the distance between the median and the party ideal point) and increases by a fairly substantial 0.175 in 2S elections (to 57.5% of the distance between median and party ideal point). The natural consequence of this divergence in party candidates is that election outcomes are more extreme in 2S elections than in 1S elections (column 3).

The effect of primary elections disappears, however, when feedback is introduced, as none of the treatment effects in columns (4), (5), or (6) are statistically significant. Comparing the intercepts with and without feedback suggests that this is because candidates in 1S elections take more extreme positions once feedback is introduced. Indeed, in 1S elections the average party candidate’s position is 0.407 without feedback and increases to 0.487 in elections with feedback. In 2S elections, feedback appears to have the opposite effect with average positions starting at 0.581 without feedback and decreasing to 0.532 with feedback. The effect of primaries on party candidate divergence thus disappears as the result of countervailing effects of feedback across institutions. The persistence of candidate divergence in a one-dimensional spatial setting is surprising given that previous experiments find a consistent tendency for candidates to converge to the median voter’s position (Collier et al. 1987, McKelvey and Ordeshook 1985, Morton 1993) or for outcomes to converge to the Condorcet winner (Fiorina and Plott 1978, McKelvey and Ordeshook 1982, Palfrey 2006). The difference may have to do with the fact that candidates are policy-motivated in my experiment rather than office-motivated in most previous experiments.

Looking only at average positions obscures the effects of primary elections on other aspects of the distribution of candidate positions. Although the effect of primaries on average positions in Experiment 1 is limited to elections without feedback, I find that primaries cause candidate positions to become more tightly centered around the mean—that is, less dispersed. Figure 4 plots the standard deviation of candidate and winning positions over the
Figure 4: Dispersion of positions and outcomes in Experiment 1
course of the experiment. The graphs reveal two interesting patterns in candidate dispersion. First, we see dispersion decreasing steadily over time. Thus, positions converge not to the median voter’s position, but to the mean position in both 1S and 2S elections. Second, we observe a clear effect of primary elections on dispersion. Variation in candidate positions and in the positions of winning candidates is consistently lower in 2S elections than in 1S elections. Primary elections appear to reinforce candidate polarization.

Voting in Primaries

The sample dynamics and analysis of candidate positions suggest that, rather than causing or exacerbating polarization, primaries instead help to maintain polarization by playing a role in the selection of candidates, weeding out party candidates who are either too extreme or too moderate. In this section, I examine voting behavior in Experiment 1’s primaries by assessing the extent to which primary voters prefer moderates or extremists and by characterizing the behavioral rule that best fits the experimental data.

Table 3 shows how the percentage of votes for the moderate varies by the positions of the candidates. Each cell shows the percentage of votes cast for the moderate candidate (the candidate closer to the median voter) for a given range of candidate positions. When the moderate candidate’s position is close to the median voter (top row), voters overwhelmingly choose the extremist, especially when both candidates are close to the median voter (only 10% prefer the moderate in the top left cell). As the extremist becomes more extreme (moving from left to right), voting for the moderate increases. However, when faced with a moderate candidate close to the median voter and an extreme candidate close to their own party’s ideal point, voters still tend to select the extremist (only 41% of votes are for the moderate in the top right cell). As the moderate candidate’s position becomes more extreme (moving from top to bottom), voting for the moderate generally increases but never exceeds 57%. The data from Experiment 1 suggests that voters fail to generate incentives for moderation consistent with the Nash predictions. If anything, their behavior favors extremists over moderates, but
Table 3: Votes for moderate by candidates’ positions in Experiment 1

<table>
<thead>
<tr>
<th>Extremist’s Position</th>
<th>[0, .2)</th>
<th>[.2, .4)</th>
<th>[.4, .6)</th>
<th>[.6, .8)</th>
<th>[.8, 1]</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, .2)</td>
<td>10%</td>
<td>23%</td>
<td>27%</td>
<td>36%</td>
<td>41%</td>
<td>31%</td>
</tr>
<tr>
<td>(10)</td>
<td>(115)</td>
<td>(165)</td>
<td>(135)</td>
<td>(100)</td>
<td>(100)</td>
<td>(525)</td>
</tr>
<tr>
<td>[.2, .4)</td>
<td>36%</td>
<td>43%</td>
<td>48%</td>
<td>53%</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>(100)</td>
<td>(195)</td>
<td>(185)</td>
<td>(95)</td>
<td>(575)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate’s Position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[.4, .6)</td>
<td>47%</td>
<td>57%</td>
<td>50%</td>
<td>53%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(85)</td>
<td>(190)</td>
<td>(145)</td>
<td>(420)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[.6, .8)</td>
<td></td>
<td></td>
<td></td>
<td>55%</td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>(20)</td>
<td>(110)</td>
<td>(130)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[.8, 1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>(30)</td>
<td>(30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10%</td>
<td>29%</td>
<td>38%</td>
<td>48%</td>
<td>49%</td>
<td>43%</td>
</tr>
<tr>
<td>(10)</td>
<td>(215)</td>
<td>(445)</td>
<td>(530)</td>
<td>480</td>
<td>1,680</td>
<td></td>
</tr>
</tbody>
</table>

This tendency is weak and does not generate the kind of strong incentives for divergence that would be encouraged by fully myopic voters.

Table 4 aggregates the voting classifications for each group of 10 elections for three different behavioral rules. The first row shows the percentage of votes for the moderate candidate. Notice that less than half of votes cast favor the moderate candidate, 37% in the first 10 elections without feedback, increasing slightly to 44 – 46% in elections with feedback. The slight increase in voting for moderates appears to lend some support for the theoretical framework, as the change in positioning behavior when feedback is introduced is consistent with the change in voting behavior. Without feedback, more votes are cast for extremists than moderates (63% versus 37%); if candidates expected this, then their best responses would have been to take more extreme positions, which is consistent with the effect.
Table 4: Primary voting behavior in Experiment 1

<table>
<thead>
<tr>
<th>Voting rule</th>
<th>Elections</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-10</td>
<td>11-20</td>
<td>21-30</td>
<td>31-40</td>
</tr>
<tr>
<td>Closer to median voter</td>
<td>37%</td>
<td>44%</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td>(525)</td>
<td>(575)</td>
<td>(630)</td>
<td>(635)</td>
</tr>
<tr>
<td>Closer to midpoint</td>
<td>67%</td>
<td>65%</td>
<td>64%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>(645)</td>
<td>(640)</td>
<td>(680)</td>
<td>(655)</td>
</tr>
<tr>
<td>Closer to own promise</td>
<td>45%</td>
<td>76%</td>
<td>75%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>(648)</td>
<td>(644)</td>
<td>(671)</td>
<td>(652)</td>
</tr>
</tbody>
</table>

of primaries in election 1-10. When feedback is introduced, there is an uptick in voting for moderate candidates, which would lead candidates to expect less extreme opponents and hence to moderate their own behavior.

Because primary voters do not express clear, unconditional preferences for either moderate or extremist party candidates, I investigated two additional behavioral rules. The first rule is a simple “midpoint” strategy where voters select the candidate closest to the midpoint between the median voter and their party’s position. The second row of Table 4 shows that this rule is a better description of behavior than voting for the moderate (or the extremist), as roughly two-thirds of votes (overall 65%) are consistent with this rule (compared to 43% consistent with voting for moderates and 57% voting for extremists). As discussed in the theoretical analysis, this voting rule has the effect of reducing the variance in party candidates.

The second alternative voting rule assumes that voters behave as if they have heterogeneous “belief-induced ideal points” $v^*_i$ (as described in the theoretical framework), and this rule appears to be the most consistent with the data. This rules assumes that each voter has an individual belief that a candidate located at $v^*_i$ maximizes their expected utility and therefore votes for the candidate closest to $v^*_i$. In Experiment 1, subjects effectively express such belief-induced ideal points when they choose campaign promises at the beginning of
each election, so I use a subject’s campaign promise as a measure of their belief-induced ideal point. The last row of Table 4 shows the percentage of votes cast in favor of the candidate closest to each voter’s own campaign promise. This voting rule, which might be considered to be consistent with a form of strategic sophistication, attains the highest classification success in elections with feedback, outperforming the simple midpoint rule. By elections 31-40, 78% of votes are consistent with voting for the candidate closest to the belief-induced ideal point (one’s own promise earlier in the election), compared to 46% for moderates, 54% for extremists, and 65% for the midpoint. Because campaign promises and belief-induced ideal points diverge from the median voter’s position, this voting rule has the effect of reducing variance in candidate positions and reinforcing candidate polarization.

Experiment 2

Procedures

I designed a second version of the experiment with two goals: to increase the salience of the candidate positioning decisions and to create an experimental environment that more closely matches the theoretical analysis of best responses to out-of-equilibrium beliefs. Increased salience was achieved by assigning subjects to fixed roles in a way that is more standard in incentivized laboratory experiments. Instead of choosing positions in each round before candidates are selected (as in Experiment 1), subjects are randomly assigned to roles as candidates and voters before the first election and retain their roles throughout the experiment. In the 1S condition of Experiment 2, all subjects are candidates and are randomly matched in pairs for each election (one left candidate against one right candidate). In the 2S condition of Experiment 2, groups of 3 (two candidates and one voter) are matched against each other, so each play of the game involves 6 subjects. There are 30 elections in Part 1, all with feedback, so Part 1 of Experiment 2 is a close analogue to Part 2 of Experiment 1 (the 30 elections with feedback). For experiment 2, I conducted two sessions with 1S elections (36
subjects) and three sessions with 2S elections (48 subjects) at the Pittsburgh Experimental Economics Laboratory in February 2016.

Later elections in Experiment 2 (elections 31-50) are designed to elicit candidates’ choices in an experimental setting closer to the theoretical analysis of best responses to out-of-equilibrium beliefs. Rather than allowing beliefs about opposing candidates to arise endogenously as uncontrolled, subjective beliefs, I rely on experimental control over the distribution of candidates. More specifically, in Part 2 of the 1S condition (20 elections), opposing candidates’ positions are not chosen by another human subject but are instead drawn randomly from a uniform distribution (over the positions between the median voter’s ideal point and the opposing party’s ideal point). Thus, I achieve control over the beliefs about the distribution of opposing candidates by controlling the positions of the opposing candidates themselves.

The procedure in the 2S condition is somewhat different to allow human voters to select candidates within each primary. The aim was to create a setup in which the initial distribution of candidates within each party is identical to the 1S election but where the distribution of the candidates in the general election depends on the behavior of primary voters. This setup closely matches the theoretical analysis while at the same time allowing the effect of primaries to arise endogenously from subjects’ behavior. In elections 31-40, all subjects act as voters and are paired against one voter from the other party. The voters simultaneously choose between two random candidates from a uniform distribution on their own side of the policy space, and the outcome of each election is the candidate closest to the median voter’s position. In elections 41-50, all subjects then act as candidates and face an opposing (computer) candidate with a position drawn randomly from the results of the previous set of elections (31-40). This design allows voting behavior to arise endogenously (in elections 31-40) and then holds it constant in subsequent elections (41-50) to preclude changes in voting behavior that might result from strategic interaction; this setup also removes any
potential for intra-party competition and renders beliefs about opposing primary voters’ behavior as the only factor relevant to the positioning decision.

Results

Figure 5 shows the average positions over time in Experiment 2, plotted separately for 1S and 2S elections. In contrast to Experiment 1, primaries in Experiment 2 cause candidates to take more extreme positions than they do in 1S elections. The top panel of Figure 5 suggests that this effect is modest but persistent over time. Similar to Experiment 1, I find that positions consistently diverge from the median voter’s position in all 30 elections regardless of the election format. In the first five rounds, the average normalized position in 1S elections is .376 compared to .531 in 2S elections. In the last five rounds, the average in 1S elections diminishes to .251 compare to .353 in 2S elections. The regression analysis in Table 5 provides more precise estimates of the effect of primaries while controlling for experience. Primary elections have a significant effect on the divergence of all candidates’ positions from the median voter (column 1), which then translates to a greater divergence in party candidates’ positions (column 2), and election outcomes (column 3). Every candidate decision is consequential, yet increasing the salience of candidates’ decisions is not sufficient to generate full convergence to the median voter’s ideal point even though candidates’ positions gradually become more moderate over time.

Turning now to the elections against random opponents’ positions, I find that behavior against random candidate positions is no different than behavior against human players. In 1S elections, the mean normalized position is .329 against human candidates and .328 against randomly drawn positions. In 2S elections, the difference in candidate positions is statistically significant when all rounds are compared (.417 against humans versus .359 against random positions, \( p < .01 \)), but this difference disappears when accounting for learning by using only the last 10 elections against human players for the comparison (.344 against humans versus .359 against random positions, \( p = .53 \)). In addition, there is no difference
Figure 5: Average positions and outcomes in Experiment 2
Table 5: Regression analysis of positions in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.088**</td>
<td>0.056**</td>
<td>0.057**</td>
</tr>
<tr>
<td>Experience</td>
<td>-0.006**</td>
<td>-0.006**</td>
<td>-0.004**</td>
</tr>
<tr>
<td>Constant</td>
<td>0.417**</td>
<td>0.415**</td>
<td>0.253**</td>
</tr>
<tr>
<td>Observations</td>
<td>2040</td>
<td>1560</td>
<td>780</td>
</tr>
<tr>
<td>R^2</td>
<td>0.0391</td>
<td>0.0338</td>
<td>0.0400</td>
</tr>
</tbody>
</table>

* \( .05 \) ** \( .01 \)

in strategic voting when selecting between random positions and positions chosen by human players, though the overall rate of voting for moderate candidates is higher in Experiment 2 than it was in Experiment 1 (64% of votes for moderates against human players and 66% against random positions, \( p = .59 \)). These results suggest that candidates in Experiment 2 choose positions as if their opponents choose their positions randomly, providing some additional support for the behavioral theory.

**Discussion**

Do primaries cause polarization? In this paper, I focus on one specific feature of primary elections that has the potential to cause increased polarization: the possibility that voters behave myopically rather than strategically. Whereas standard game theory predicts the complete absence of polarization and the irrelevance of primaries, behavioral game theory accounts for the substantial candidate divergence observed in the experiment and, moreover,
identifies the close connection between strategic expectations and polarization. A candidate’s optimal position depends on what that candidate expects the opposing party’s voters to do, and the primary election is a tool that partisan voters use to select candidates.

Experimentally, I find that primaries have a modest effect on polarization in the absence of feedback in Experiment 1 but consistently in Experiment 2. Voters appear to be sophisticated enough to use the primary process to weed out candidates both too close and too far from the general election median voter’s position. Thus, voters in the lab seem to recognize the tension between centrist policies that yield few policy benefits and extreme positions that are unlikely to win the general election, resolving the trade-off by generally splitting the difference. In a sense, primary elections generate ideological “purity”—not because voters care about ideological purity per se, but because nominees selected by primaries reflect a majority of partisan voters’ beliefs about optimal positioning (thereby correcting perceived mistakes that party candidates might make when left to their own devices).

The experimental findings that candidates diverge generally match what social scientists know about primary elections from observational research, and the correspondence between these key features of the laboratory experiment and the real world suggests several ways in which my analysis enhances our understanding of partisan electoral politics. First, it may be that partisan primary voters are politically and strategically more sophisticated than is typically presumed in popular discussions of primaries (the caricature of partisan zealots seeking ideological purity). Second, the results suggest that partisan polarization may have less to do with the specific features of primary elections or with the preferences of primary voters than with general political forces that attract the activists, donors, and candidates for office who shape the party’s ideological positions. That is, the fact that partisan voters have preferences that diverge from the general election median voter is not sufficient to cause a substantial increase in polarization. Closed primaries may exacerbate polarization, but they are unlikely to be the main cause. Third, strategic expectations about increasing polarization can, to some extent, be self-fulfilling. If party A’s candidates are policy-motivated and
expect party B’s candidates to be increasingly more extreme, then party A’s candidates will
take more extreme positions to maximize their expected policy benefits. The only problem
with this argument is that to the extent that general elections favor centrist candidates,
the incentives for moderation remain theoretically powerful. But in practice, there may be
sufficient heterogeneity in candidate positions to weaken the influence of the median voter.
Although parties are long-lasting, their decentralized organization may be enough to prevent
equilibrium convergence.

References

Adams, James and Samuel Merrill. 2008. “Candidate and party strategies in two-stage
elections beginning with a primary.” American Journal of Political Science 52(2):344–
359.

Adams, James and Samuel Merrill. 2014. “Candidates’ policy strategies in primary elections:
does strategic voting by the primary electorate matter?” Public choice.

Ansolabehere, Stephen, James M Snyder, Jr. and Charles Stewart, III. 2001. “Candidate
159.

Aronson, Peter H. and Peter C. Ordeshook. 1972. Spatial Strategies for Sequential Elec-
tions. In Probability Models of Collective Decision Making, ed. Richard G. Niemi and

Aumann, Robert and Adam Brandenburger. 1995. “Epistemic conditions for Nash equilib-


Appendix: Instructions for Two-Stage Elections with Fixed Matching

Instructions

General Information

This is an experiment on the economics of elections. The National Science Foundation and the University of Pittsburgh have provided funds for this research.

You will be paid in cash for your participation, and the exact amount you receive will be determined during the experiment and will depend partly on your decisions, partly on the decisions of others, and partly on chance. You will be paid your earnings privately, meaning that no other participant will find out how much you earn. These earnings will be paid to you at the end of the experiment along with the $7 participation payment.

Pay attention and follow the instructions closely, as we will explain how you will earn money and how your earnings will depend on the choices that you make. Each participant has a printed copy of these instructions, and you may refer to them at any time.

If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Also, please ensure that any phones or electronic devices are turned off and put away. Participants intentionally violating the rules will be asked to leave and may not be paid.

Parts and Elections

This experiment consists of several parts. Each part consists of a series of elections, and we will explain the instructions for each part before beginning that part.

We will randomly select one election to count for payment from the entire session. Each election is equally likely to be selected. The points you receive from that election will be used to calculate your payment for the experiment, and points will be converted to cash at the rate of $1 for every 10 points. You should think of each election as a separate decision task.

Before we begin, we will randomly divide you into two groups of seven participants. This random assignment will take place once so that the members of your group will be the same in every round.
Appendix: Instructions for Two-Stage Elections with Fixed Matching

Part 1

There will be 10 elections in Part 1, and each election consists of three stages: a campaign stage and two voting stages.

Campaign Stage

In the campaign stage, you will choose a whole number from 1 to 200. This number is your “campaign promise” and you can think of it as a position or stance on a particular policy issue that both voters and candidates care about. If you are selected as a candidate and you win the election, then this number will determine the payoffs for each voter and candidate (as we will explain later).

Once all of the participants have chosen a campaign promise, the computer will then select two members of each group at random to be candidates (all members of the group are equally likely to be selected), and then we will move to the voting stages. Note that even though only two members of your group will be selected as candidates, you should choose your campaign message as if you were actually selected as a candidate.

First Voting Stage

The members of each group who are not selected as candidates will be the voters. Thus, in each group there will be 2 candidates and 5 voters.

In the first voting stage, each group votes to determine which of the group’s two candidates will be a candidate in the second voting stage. The candidate who receives the most votes will move on to the second voting stage. Only members of your group will be voting on the candidates from your group in the first voting stage. The other group will be voting on their own two candidates at the same time.

Second Voting Stage

The winners of each group’s first vote will then be the two candidates in the second voting stage. That is, there is one candidate from each group in the second stage. All voters from both groups will vote in the second stage. The candidate who receives the most votes wins the election. In addition, there will be one “computer voter” in the second stage of voting. The computer voter is not a member of either group but is like a robot programmed to always vote for the candidate whose campaign promise gives it the higher payoff value (the promise closest to its own favorite position, as described below). If both candidates in the second stage offer the computer voter the same payoff, then the computer voter will cast its vote randomly between the two candidates (with votes for each candidate equally likely).
Appendix: Instructions for Two-Stage Elections with Fixed Matching

Payoffs

In each round, you will be assigned a “favorite position” and you will earn points based on how close the winning candidate’s campaign promise is to your favorite position.

The closer the winning campaign promise is to your favorite position, the more points you will earn. Specifically, we will compute the absolute difference between the winning campaign promise and your favorite position and then subtract this amount from 200. This is described by the following formula:

\[
\text{Points} = 200 - |\text{Winning campaign promise} - \text{Your favorite position}|
\]

For example, if your favorite position is 57 and the campaign promise of the candidate who wins the second election is 27, then your points from that election are 200 - |57 - 27| = 200 – 30 = 170 points. Of course, this is just one example. Note also that candidates and voters (including the computer voter) all earn points according to the same formula; candidates do not earn extra points for winning. Remember that we will pay you $1 for every 10 points you earn (rounded to the nearest quarter).

In every election, each group will have a different favorite position. Within groups, every member’s favorite position will be the same. For example, if your group’s favorite position is 50 and the other group’s favorite position is 150, then everyone in your group has a favorite position of 50 while everyone in the other group has a favorite position of 150. The computer voter will always have a position that is somewhere between the two groups. Everyone will find out the values of these favorite positions at the beginning of each election.

Sequence of Decisions

In Part 1 you will make your decisions for all of the elections in each stage separately before moving on to the next stage. In other words, first you will choose your campaign promises for all elections before moving to the voting stages. Second, you will cast your votes in all first voting stages for which you are a voter. Finally, you will cast your votes in all second voting stages for which you are a voter. Note that you will not receive any feedback about results of the elections from Part 1.

Summary

1. Before any of the elections, you are randomly divided into two groups. These groups will remain the same throughout the experiment.

2. In every round, you will find out the favorite position of your group, the other group, and the computer voter.
Appendix: Instructions for Two-Stage Elections with Fixed Matching

3. In the campaign stage, you choose a number from 1-200 that serves as your campaign promise. If you are elected, this campaign promise determines everyone’s payoff.

4. We randomly determine two members of each group to be candidates and the other members to be voters.

5. In the first voting stage, each group simultaneously selects which of its candidates to put forward in the second voting stage.

6. In the second voting stage, there is one candidate from each group. In addition, there is a computer voter that will vote for the candidate whose campaign promise is closest to its own favorite position. The campaign promise of the candidate who wins the second voting stage determines everyone’s payoff for that election.

7. Your payoff is:

   \[ \text{Points} = 200 - |\text{Winning campaign promise} - \text{Your favorite position}| \]

   The closer the winning campaign promise is to your favorite position, the more points you will earn.

Are there any questions? If you have a question, please raise your hand.

**Instructions Quiz**

Before we begin the experiment we would like you to answer a few questions to make sure you understand how election experiment works. You will answer these questions on your computers and will receive immediate feedback once you answer all of the questions. We will then begin the experiment when everyone has answered these questions.

1. If your favorite position is 20 and the winning candidate’s campaign promise is 90, how many points would you earn?

2. If your favorite position is 165, your campaign promise is 150, and you win the election, how many points would you earn?

3. Suppose that the results of the first stage votes are as follows. In Group L, Candidate A receives 3 votes and Candidate B receives 2 votes. In Group R, Candidate C receives 1 vote and Candidate D receives 4 votes. Which candidates will you be able to vote for in the second stage vote?

4. Suppose that the computer voter’s favorite position is 80. If Candidate A’s campaign promise is 50 and Candidate B’s campaign promise is 120, which candidate would the computer vote for?
Appendix: Instructions for Two-Stage Elections with Fixed Matching

Part 2

There are 30 elections in Part 2. The rules for elections in Part 2 are exactly the same as in Part 1. Each election will consist of a campaign stage and two voting stages. The payoffs will also be determined in the same way as before.

In Part 2 only the sequence of decisions and the feedback you receive about decisions will be different. Instead of making all of your decisions for each stage before moving to the next stage, you will make your decisions in sequence for each election separately. In other words, for the first election, you will choose your campaign promise, then cast your first stage vote, and then cast your second stage vote. After the second stage vote you will learn the results of the election immediately after the election is finished. You will then move on to the next election, beginning with the campaign promise stage, and so on.

Are there any questions? If you have a question, please raise your hand.