Primaries and Candidate Polarization: Behavioral Theory and Experimental Evidence

Jonathan Woon[*]
University of Pittsburgh

August 23, 2017[†]

Abstract

Do primary elections cause candidates to take extreme, polarized positions? Standard equilibrium analysis predicts full convergence to the median voter's position, but behavioral game theory predicts divergence when players are policy-motivated and have out-of-equilibrium beliefs. Theoretically, primary elections can cause greater extremism or moderation, depending on the beliefs candidates and voters have about their opponents. In a controlled incentivized experiment, I find that candidates diverge substantially and that primaries have little effect on average positions. Voters employ a strategy that weeds out candidates who are either too moderate or too extreme, which enhances ideological purity without exacerbating polarization. The analysis highlights the importance of behavioral assumptions in understanding the effects of electoral institutions.

(Word count: $9,098$)

[*] 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, Danielle Thomsen, Alan Wiseman, participants at the Yale CSAP American Politics Conference, seminar participants at Washington University in St. Louis, University of Oxford (Nuffield College CESS), IC3JM (Juan March Institute), and 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)[1]

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 for this phenomenon on partisan primary elections, typically employing 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.

Political scientists have tested this argument, finding that while there is some evidence to suggest that primary elections promote extremism, the empirical record is generally 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). These findings seem puzzling in light of the basic theory of representation at the heart of this literature: that the preferences of primary electorates should affect the preferences of party candidates.

[1] Charles E. Schumer, "Adopt the Open Primary," *New York Times*, July 21, 2014.

How, then, should we understand the causal relationship between primary elections and candidate positioning? I examine the connection, both theoretically and experimentally, by comparing elections with and without primaries while holding other features of the electoral environment constant, including 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.[2]

The theory that I develop suggests a more nuanced relationship between primaries and polarization than portrayed in the existing literature. I show that primaries can cause polarization *or* moderation, depending on candidates' beliefs about opposing voters' strategic behavior—even when preferences are held constant. To generate this insight, I rely on ideas from behavioral game theory, which retains much of the theoretical apparatus from standard game theory while allowing for key departures (Camerer 2003). Specifically, I allow players to have "incorrect" or "non-equilibrium" beliefs about others' actions (Crawford, Costa-Gomes and Iriberri 2013), but assume they are nevertheless strategic in the sense that they best respond to what they *think* other players do (Camerer, Ho and Chong 2004, Nagel 1995, Stahl and Wilson 1995). The analysis demonstrates that changes in preferences alone are insufficient to cause polarization. Instead, beliefs and expectations about the strategic behavior of others play important roles in conditioning the effect of institutions.

I turn to the laboratory and conduct a series of experiments to test the effects of primaries on candidate positions. The chief advantage of the laboratory for theory testing is control (Aldrich and Lupia 2011, Falk and Heckman 2009, Morton and Williams 2010), so we can be confident that the observed behavior occurs under the same conditions specified by the theoretical model. Indeed, subjects face the same key trade-off in the experiment as the actors do in the model between increasing the favorability of winning outcomes versus increasing the probability of winning. In the lab, theoretically-relevant quantities of interest

[2] Such considerations include 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).

that are difficult to measure using observational data with any accuracy or without strong assumptions (in particular, preferences and positions) are also known exactly. Furthermore, experimental manipulations permit tests of mechanisms not possible using observational data. Thus, laboratory experiments are ideal for theory testing given their high internal validity.[3]

The key finding from the experiment is that primaries appear to cause a kind of ideological purity rather than greater extremism. I find that subjects take positions that diverge significantly from the median voter's position, regardless of whether or not there is a primary. This finding lends support for the behavioral theory. However, the extent to which primaries cause polarization is limited. Greater polarization only occurs when there is no feedback such that candidates cannot learn about the behavior of others, and this polarization happens because voters tend to select extremists over moderates, even though candidate positions do not vary with the election format. More precisely, the analysis reveals that voters 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 that diverges from the median voter. Hence, greater ideological purity reinforces, rather than exacerbates, polarization.

[3] The main question of interest for theory testing, as Aldrich and Lupia (2011, 90) put it, is "Will people who are in the situations you describe in your model act as you predict?" Also see Dickson (2011), Palfrey (2006), and especially, Morton and Williams (2010). While the question of external validity ("to what extent can we generalize from a particular sample?") is an enduring source of controversy in political science, Falk and Heckman (2009) argue in their insightful defense of the value of lab experiments in social science that "Behavior in the laboratory is reliable and real: Participants in the lab are real human beings who perceive their behavior as relevant, experience real emotions, and take decisions with real economic consequences" (536). Indeed, there are many precedents for testing theories of elite behavior using laboratory experiments (e.g., Aragones and Palfrey 2007, Frechette, Kagel and Lehrer 2003, Morton 1993). Moreover, Druckman and Kam (2011) note that there is nothing *inherently* problematic with using student samples, and there is little evidence to suggest that using undergraduates as stand-ins for elites biases the results in any particular direction (see Morton and Williams 2010, 343–347). For example, Potters and van Winden (2000) find significant, but small, differences between students and lobbyists, Fatas, Neugebauer and Tamborero (2007) find elites do not fit prospect theory as well as students, while studies by Belot, Duch and Miller (2015), Cooper et al. (1999), and Mintz, Redd and Vedlitz (2006) suggest that student samples provide a *lower bound* to departures from rational decisionmaking.

Related Literature

My analysis follows a long tradition of using spatial voting models to understand elections. 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).[4] These models also assume that general election outcomes are probabilistic, which is theoretically consequential because the mechanism they rely on to produce divergence 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 the literature whether polarization can be traced to any distinctive features of primaries per se, as electoral institutions. By explicitly comparing institutions, my analysis speaks directly to the connection between primaries and polarization.

Existing theoretical models of two-stage elections also typically maintain the assumption that all political actors, candidates as well as voters, are strategic and forward-looking (e.g., Owen and Grofman 2006). Several models consider the issue of raiding and cross-over voting in open primaries (Cho and Kang 2014, Chen and Yang 2002, Oak 2006), which requires a fairly high degree of strategic sophistication, but this kind of behavior is outside the scope of my analysis. My results also differ 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. In contrast to the preponderance of existing formal models, I take a behavioral (i.e., bounded rationality) approach advocated by Simon (1955), Ostrom (1998) and others. I do so by explicitly allowing for sincere or myopic behavior as well as subjective beliefs that are inconsistent with observed behavior.

[4] An 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.

This paper is also related to 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 and, 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). An exception is when candidates are ideological and voting is probabilistic (Morton 1993). The other related literature, on strategic voting, generally finds little (or at best, mixed) evidence for voter sophistication in the early stages of a multi-stage voting agenda or election contest (Cherry and Kroll 2003, Eckel and Holt 1989, Herzberg and Wilson 1988, McCuen and Morton 2010, Plott and Levine 1978, Van der Straeten et al. 2010).[5] Taken together, these previous studies raise doubts that voters will be highly strategic (even if candidates are), calling into question theories predicated on voter rationality and strategic sophistication.

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 three ideal points: θ_L for members of Party L, θ_R

[5] An exception is Smirnov (2009), who studies endogenous agendas and finds behavior consistent with sophisticated expected utility maximization. There is stronger experimental evidence for other kinds strategic voting, however, such as coordinating on a less-preferred candidate in multi-candidate contests (Rietz 2008), and in incomplete information pivotal voter settings (e.g., Battaglini, Morton and Palfrey 2010).

for members of Party R, and θ_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, $u_i(w) = K - |w - \theta_i|$, for $i \in \{L, M, R\}$ and some constant 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 θ_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 θ_R chooses $c_R \in \{c_{R1}, c_{R2}\}$ in the Party R primary. In the general election, the median voter with ideal point θ_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 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 micalculating 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 also 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 thus establishes a critical link between the effect of primaries and candidates' beliefs about opponents' primary voting behavior.

Candidate equilibrium with fully strategic voters

Equilibrium theory 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 θ_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 similar to 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 θ_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) the 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."[6] 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 while the equilibrium of the one-stage election game remains the same (full convergence, since there are no primary voters).

[6] While the overall level of voter "rationality" remains an ongoing subject of 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 2012). A theory of elections with boundedly rational, behavioral voters is also worked out by Bendor et al. (2011).

In any equlibrium 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.[7] 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, $E[w] = \theta_M$. 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 weakly more polarized in two-stage elections than in one-stage elections.*

[7] 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. That is, if candidate j chooses the platform c_j, then candidate i must believe with certainty that c_j must really be j's position. However, this 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.[8]

To model this, let candidate i's beliefs about the positions of candidates from the opposing party $j \neq i$ be given by the cumulative distribution $F(c_j)$. Importantly, these beliefs need not be accurate. For instance, if j's true position is $c_j = 0$, candidate i might believe that c_j is uniformly distributed between -1 and 1. We can think of the distribution $F(c_j)$ as representing *subjective beliefs* that will typically not satisfy the equilibrium consistency requirement.[9]

By relaxing the standard equilibrium assumption of belief consistency, an otherwise expected utility maximizing candidate will choose a position that diverges from the median voter's ideal point. The reasoning is as follows. If a candidate believes there is *some* possibility that the opposing candidate's position diverges from the median voter, then it cannot be optimal for a policy-motivated candidate to choose a platform exactly at the median voter's ideal point. Instead, the candidate will choose a position that trades off some probability of

[8] While level-k models are a subset of the class of models that assume out-of-equilibrium beliefs, my theory does not rely on different levels of sophistication or reasoning as modeled explicitly in the level-k framework.

[9] I assume that the PDF $f(c_j)$ has full support over the interval between median voter θ_M and the opposing party θ_j. The distribution $F(c_j)$ can also be interpreted as an objective probability distribution if candidates' choices are noisy and $F(c_j)$ reflects the true distribution of candidate positions.

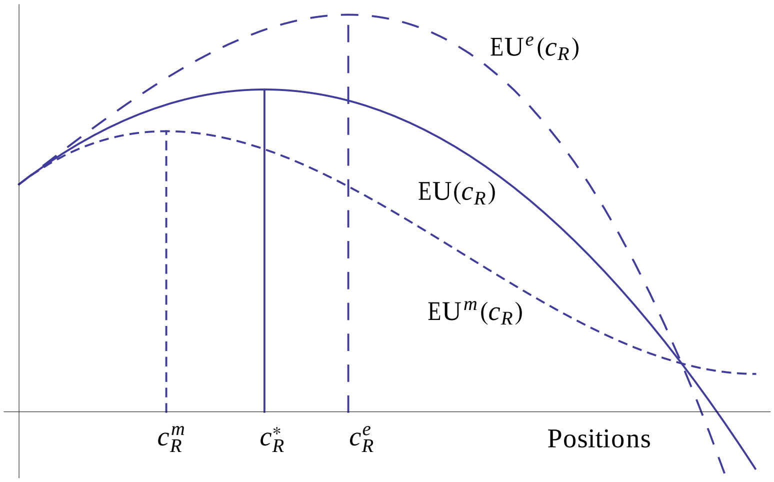


Figure 1: Comparison of candidate expected utility in one-stage and two-stage elections as a function of non-equilibrium beliefs and opponents' primary voting

winning against potential policy gains obtained from choosing a position closer to his or her own ideal point. When a candidate is more likely to expect her opponent to be extreme (i.e., when $F(c_j)$ puts more weight on extreme positions), then she herself will take a position with greater divergence from the median voter in response. To illustrate this concretely, 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 this trade-off is $c_R^* = \frac{1}{3}$. This is illustrated by the solid line in Figure 1 showing the expected utility function $EU(c_R)$.

Next, I consider how these beliefs about opposing candidates' positions interact with the election format. The main result is that the effect of primaries will depend on the candidates' beliefs about the opposing party's primary voters. The baseline for comparison is a one-stage election with opponents drawn from the belief distribution $F(c_j)$. For the purposes of exposition, suppose that $F(c_j)$ is uniform as in the example just given and as shown in the left side of Figure 2, so the candidate's best position is the one that maximizes the expected utility function shown by the solid line in Figure 1.

In a two-stage election, it is not the original distribution of candidates $F(c_j)$ that matters, but beliefs about which candidate will emerge from the primary election. Let

$G(c_j)$ denote this latter set of beliefs about the candidate selected by the opposing party's primary—the candidate that i expects to face in the general election. Intuitively, we can think of the primary election as a selection mechanism or filtering process that affects whether a party's candidate is systematically more or less extreme than the party's initial set of candidates.

More precisely, suppose that both of the opposing party's candidates are independently drawn from $F(c_j)$. Now consider how primary voting behavior affects $G(c_j)$ and, in turn, candidates' positions. If j's primary voters unconditionally select the more extreme candidate (as they would if they voted sincerely), then party j's candidate in the general election will be the more extreme of two independent draws from $F(c_j)$. This results in a distribution $G(c_j)$ that is skewed more towards j's own ideal point than $F(c_j)$, as shown by the triangular distribution in the upper-right of Figure 2 when $F(c_j)$ is uniform. When voters choose extremists, primaries generate incentives for greater *extremism* than in one-stage elections, as illustrated by the upper-dashed expected utility function $EU^e(c_R)$ in Figure 1.

The flip-side of this is that if j's primary voters select the more moderate candidate (as they would in equilibrium, they generate incentives for greater *moderation* than in one-stage elections. This is because party j's general election candidate will be the more moderate of two independent draws from $F(c_j)$, resulting in a distribution of beliefs $G(c_j)$ that is skewed more towards the median voter than $F(c_j)$, as shown by the triangular distribution on the bottom-right of Figure 2. When the probability of facing an extremist opponent is lower, a candidate must moderate their position in response, which is shown by the lower-dashed expected utility function $EU^m(c_R)$ in Figure 1.

These are not the only possibilities, as primary voters might also behave in other ways. For example, a fairly sophisticated voter might reason in the same way as a candidate and form the same beliefs $G(c_i)$ based on expectations about opposing primary voters. To generalize this a bit, suppose that a voter has a *belief-induced ideal point* c_j^* and always votes for the candidate in the primary whose position is closest to c_j^*; sometimes this will be the

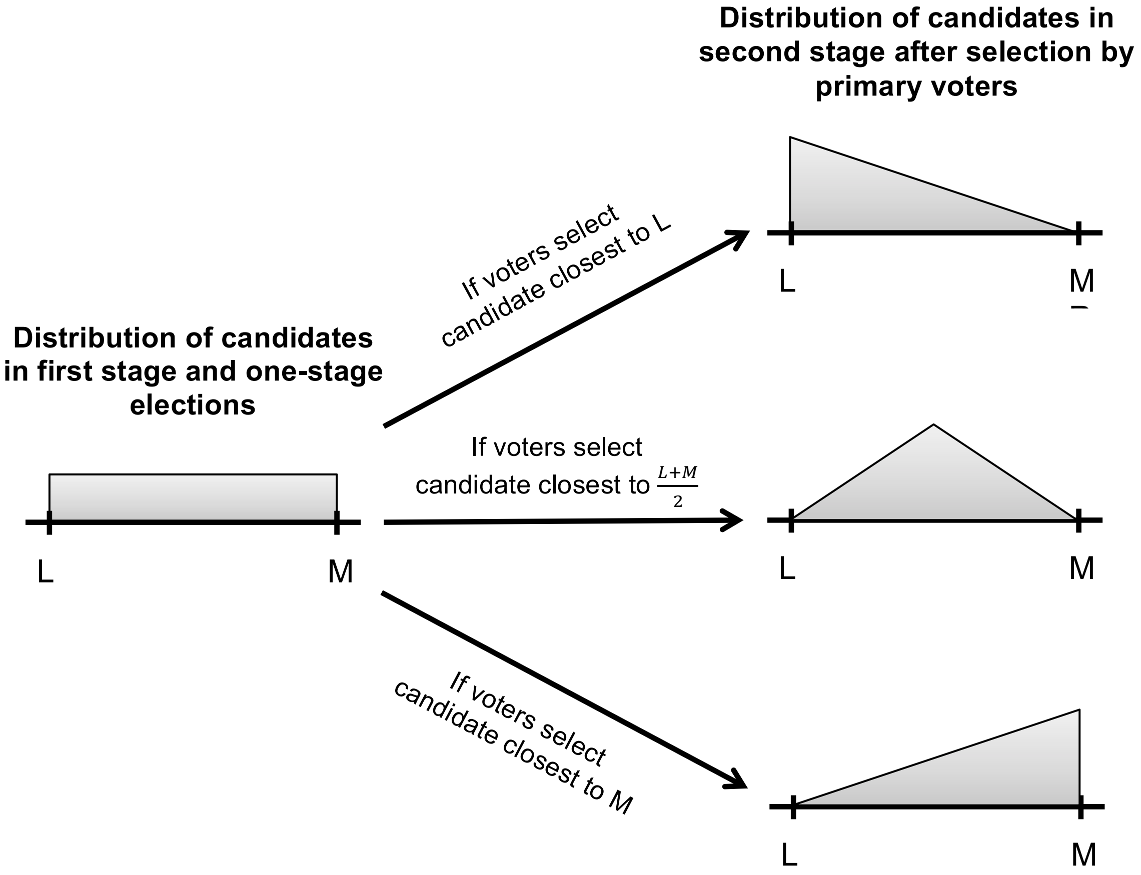

Distribution of candidates in second stage after selection by primary voters
Distribution of candidates in first stage and one-stage elections
If voters select candidate closest to L
If voters select candidate closest to $\frac{L+M}{2}$
If voters select candidate closest to M
L
M
L
M
L
M
L
M

Figure 2: Comparison of beliefs in one-stage and two-stage elections as a function of opponents' primary voting behavior

moderate and sometimes the extremist. The result is a distribution of primary candidates $G(c_j)$ that has greater mass closer to c_j^* than $F(c_j)$ does. If c_j^* happens to be the midpoint between the voter's ideal point and the median voter, $G(c_j)$ will be the symmetric triangular distribution in the middle-right of Figure 2. Note that while the mean of this distribution is the same as the original distribution $F(c_j)$, it has lower variance. Primary elections may therefore also have the effect of reinforcing a kind of ideological purity even when there is no discernible effect on average candidate positions.

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 absence of primaries and with complete information about preferences. Moreover, 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, if primary voters select moderates.

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.*

Experimental Analysis

The theoretical analysis generated a set of competing predictions about the effect of primaries as a function of alternative behavioral assumptions. If all players are fully strategic, then we should observe full convergence to the median voter's position and primaries should have no effect. If candidates are strategic but voters are not, then we should observe candidate divergence in two-stage elections but not one-stage elections. If the behavioral theory has merit and candidates have subjective beliefs about their opponents' positions, then either polarization or moderation is possible depending on voter behavior. Which set of assumptions better reflects human behavior is ultimately an empirical question, and thus, I turn to the lab.

Procedures

The experiment was conducted at ******** and involved a total of 182 participants drawn primarily from the university's undergraduate population. Each session involved 14 participants, and each subject participated in one session of either the one-stage (1S) election treatment (6 sessions) or the two-stage (2S) election treatment (7 sessions). At the beginning of each session, following standard laboratory procedures, subjects gave informed consent, the instructions were read out loud to induce public knowledge, and subjects answered a set of questions about the rules on their computers to ensure comprehension.[10] The interface was computerized and programmed using the software z-tree (Fischbacher 2007). Each session took about about an hour and a half to complete, and subjects earned an average of \$21.10 (including a \$7 show-up fee).

Subjects participate in a total of 40 elections, and the instructions emphasize that each election is to be treated as a "separate decision task." For each election, subjects are divided into two groups of seven participants, and every member of a group has the same payoff function and ideal point.[11] Throughout the experiment, the policy space is the set of integers from 1 to 200, and payoffs are given by the linear loss function $200 - |w - \theta_i|$.[12]

[10] See the Appendix for the full text of the experimental instructions. Comprehension of the instructions was high. The percentage of correct responses for individual questions ranged from 81% to 94%, and 69% answered all 4 questions correctly while only 8% missed more than one question. These figures likely underestimate the overall degree of comprehension since subjects read explanations of the correct answers before playing the game.

[11] 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. Groups were randomly reassigned between rounds in two sessions of each treatment, while the remaining sessions involved fixed groups. The method of group assignment does not affect the results, so I ingore the distinction and pool the data in the analysis.

[12] Note that with a linear loss function (in contrast to quadratic loss), every possible policy outcome

The parties' ideal points are located symmetrically from the median voter's ideal point θ_M such that $\theta_L = \theta_M - \delta$, $\theta_R = \theta_M + \delta$, and $\delta \in \{50, 75\}$. The numerical value of θ_M varied from election to election, while the exact sequence of values is identical across sessions and treatments.[13] Payoffs are denominated in "points" and converted to cash by 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 and adding the show-up fee.

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), and they know that if their campaign promise is selected as the winning position, it will affect every other subject's payoff. After subjects choose their campaign promise, the computer then 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.[14]

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. After each group

between the parties' ideal points generates an equal amount of total social welfare, making it unlikely that risk neutral, altruistic subjects will want to choose the midpoint between parties to maximize the total social monetary payoffs of both groups. However, to the extent that subjects' preferences for money exhibit risk aversion (and they expect this of other subjects), total social welfare will be maximized at the midpoint between parties, which would bias the results *toward* median convergence.

[13] To determine the sequence of values, I randomly selected the median's position, θ_m, from the integers between 51 and 150 for $\delta = 50$ and between 76 and 125 when $\delta = 75$. I varied the numerical values in order to encourage subjects to pay attention and think about their relative, rather than absolute, positions.

[14] 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.

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, which is effectively the "general election."[15] In contrast to the 2S election treatment, the 1S election treatment features only one round of voting in which every voter participates.

The median voter in the general election in both the 1S and 2S election treatments is a "computer voter" who has a distinct ideal point and, as the instructions explain to subjects (following Morton 1993), is "like a robot programmed to always vote for the candidate whose campaign promise gives it the higher payoff value." In the case of ties, the computer votes for each candidate with equal probability. The subjects are informed of the computer voter's ideal point before every election.

The 40 elections within each session are divided into two parts, with each part varying the type of feedback subjects receive. Part 1 consists of 10 elections without any feedback.[16] Part 2 consists of 30 elections with feedback provided to subjects after each 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. In the first half of the elections in each part, the left and right groups' ideal points are 100 units apart, while they are 150 units apart in the second half of the elections. Note that these two within-subjects manipulations vary ancillary assumptions (feedback and distance between ideal points) and therefore serve as robustness checks. The experimental manipulation of theoretical interest is the between-subjects manipulation of the electoral institution.

[15] To 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."

[16] The fact that the game is sequential means that 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.

Electoral Dynamics

To get a sense for the kinds of promises candidates make and whether moderates or extremists win elections, Figure 3 presents the sequence of candidate positions and outcomes for selected 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. The vertical lines show where electoral conditions change in terms of feedback and preference polarization. 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.

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 polarization between the candidates' positions increases when the underlying preference polarization increases. Along with divergence, there also appears to be substantial heterogeneity and fluctation in candidate positions.[17]

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 ±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.

[17] 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.

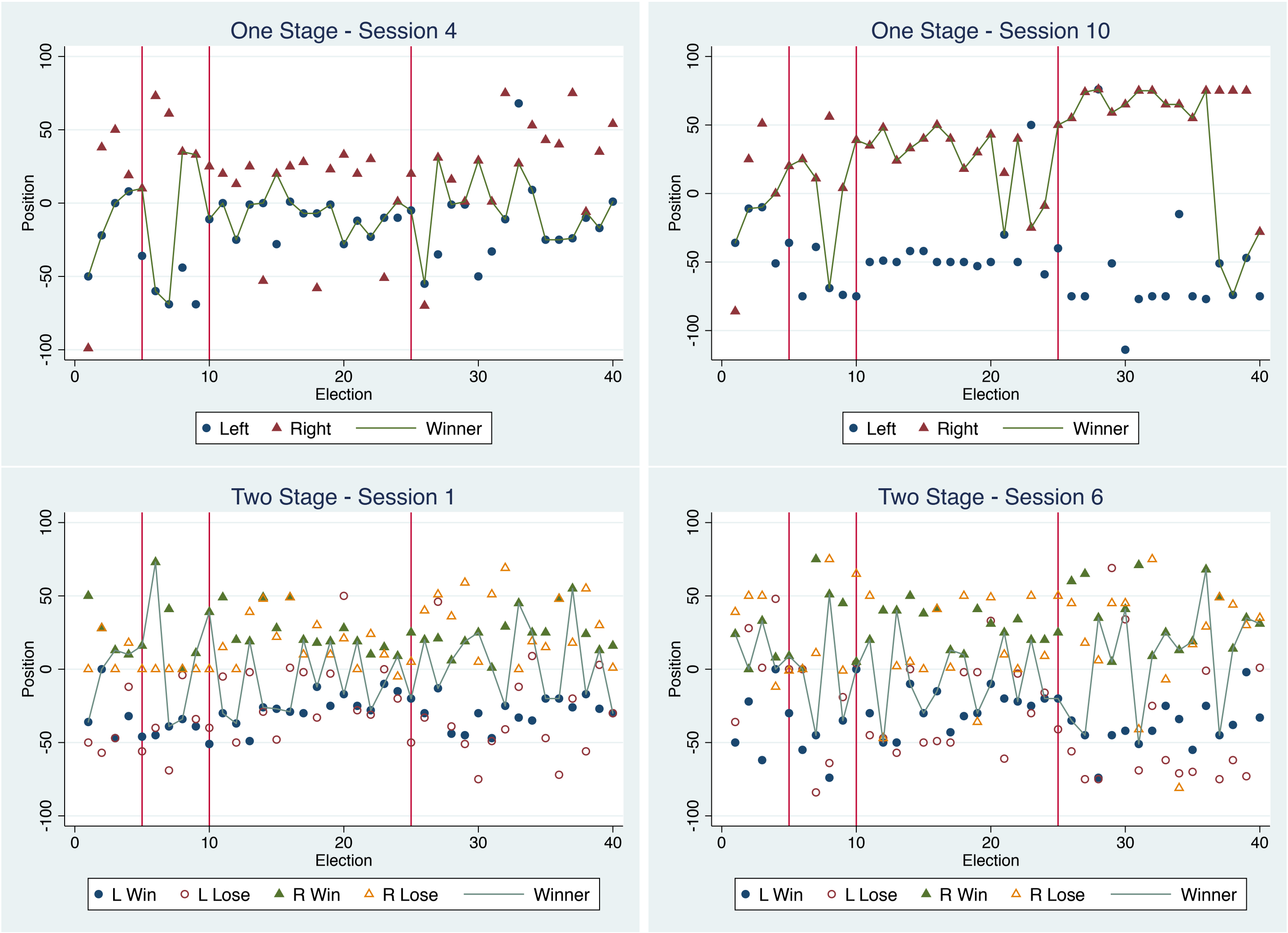


Figure 3: Sample session dynamics

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 3 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 hetergeneity 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 3 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.

Candidate Positions

Figure 4 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 4 shows that candidate positions clearly diverge from the median voter's position throughout the experiment 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 slight convergence of winning candidates to the median voter's position in elections without primaries and an increase in divergence once feedback is introduced in election 11. In elections with primaries, however, the positions of winning candidates remain polarized throughout the experiment.

Table 1 presents a series of ordinary least squares regressions to measure the effect of primaries on candidate divergence while controlling for feedback and experience. The estimates generally reinforce the visual interpretation of the data displayed in Figure 4. Positions are divergent (as measured by the intercept) and do not change over time (as the coefficients on *Experience* are small and insigificant across the models). Although primary elections have no effect on the positions chosen by all candidates (column 1), they do have a statistically significant effect on the divergence between party candidates (those standing for election in the second voting stage, column 2) 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.

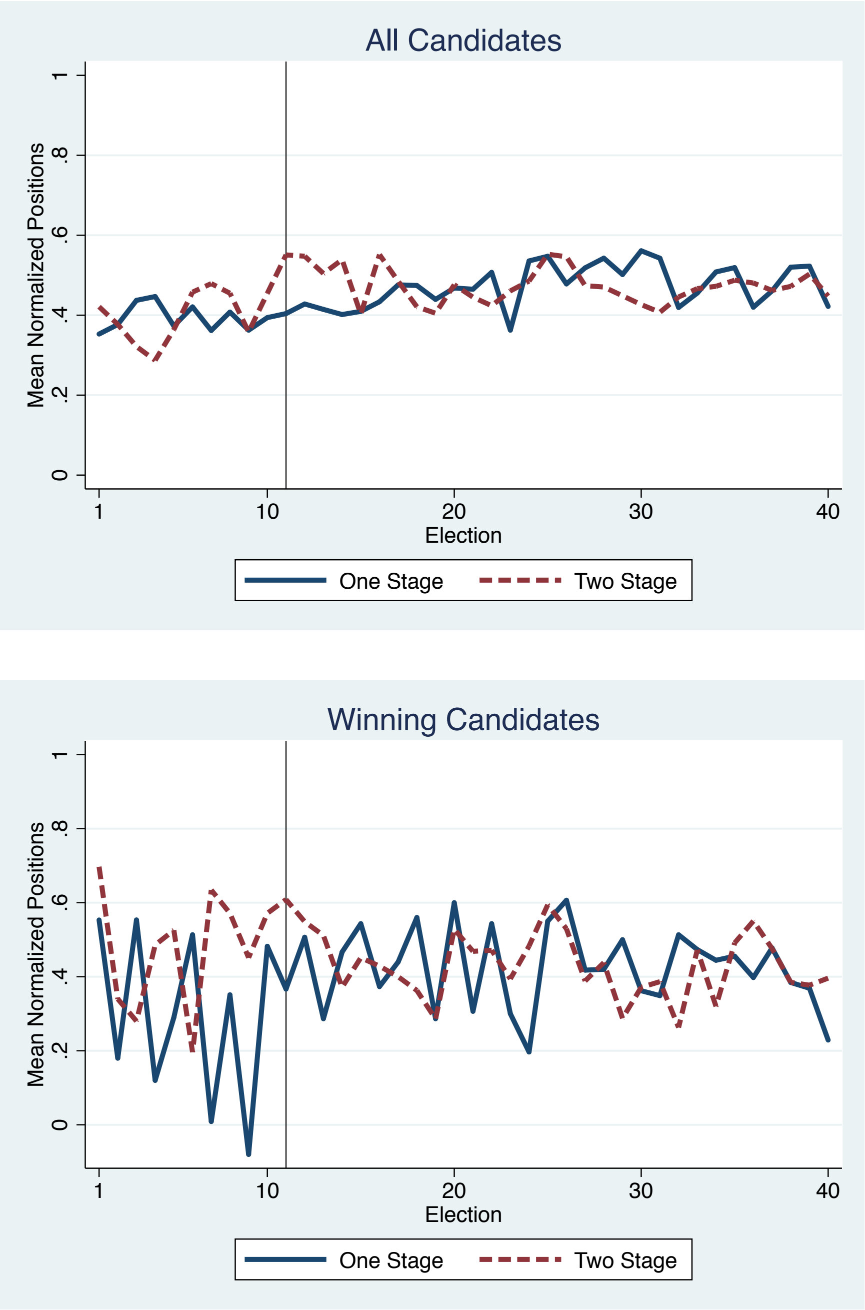


Figure 4: Average positions and outcomes

Table 1: Regression analysis of positions

	No feedback (elections 1-10)			Feedback (elections 11-40)		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Party	Winner	All	Party	Winner
Primary (2S) Elections	0.004	0.175*	0.178**	0.003	0.046	0.011
	(0.060)	(0.067)	(0.058)	(0.042)	(0.090)	(0.102)
Increased Polarization	0.081	-0.098	0.007	0.016	0.044	0.022
	(0.041)	(0.102)	(0.143)	(0.022)	(0.050)	(0.080)
Experience	-0.008	0.013	-0.007	-0.000	-0.002	-0.003
	(0.009)	(0.017)	(0.023)	(0.001)	(0.004)	(0.004)
Constant	0.387**	0.400**	0.326**	0.469**	0.499**	0.459**
	(0.058)	(0.051)	(0.062)	(0.042)	(0.103)	(0.111)
N	1820	260	130	5460	780	390
R^2	0.001	0.028	0.047	0.0001	0.004	0.004

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

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.[18]

[18] 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, but there are a number of other subtle differences between my design and previous experiments, including the use of linear instead of quadratic utility, random role assignment, the strategy method, and the varying of the numerical value of players' ideal points. Isolating the exact cause of the difference would be interesting, but is largely beyond the scope of this paper. Nevertheless, I conducted a modified version of the experiment (discussed in the Appendix) in which I increase the salience of players' decisions by assigning fixed roles. The main result that candidates diverge in both 1S and 2S elections holds up in the modified Fixed Roles Experiment. Some differences emerge, including movement towards the median over the course of the experiment and greater polarization in 2S than in 1S, though the magnitude

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 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 5 plots the standard deviation of candidate and winning positions over the course of the experiment. The graphs reveal two interesting patterns in candidate dispersion. First, we see dispersion decreasing steadily over time. Thus, because average positions remain unchanged, 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 (compare standard deviations of 0.47 in the former and 0.55 in the latter, with the difference statistically signficant according to a variance ratio test, $p < 0.01$). Primary elections therefore appear to reinforce candidate polarization.

Voting Behavior

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 primaries by assessing the extent to which primary voters prefer moderates or extremists and by determining the behavioral rule that best fits the experimental data.

Voters tend to select the more extreme candidate, but it is not an overwhelming preference. Overall, voters prefer the extremist in 57% of the elections in the data. When the moderate candidate's position is closest to the median voter (between 0 and 0.2 on the normalized scale), voters overwhemingly choose the extremist (69%), especially when both

of the effect is modest.

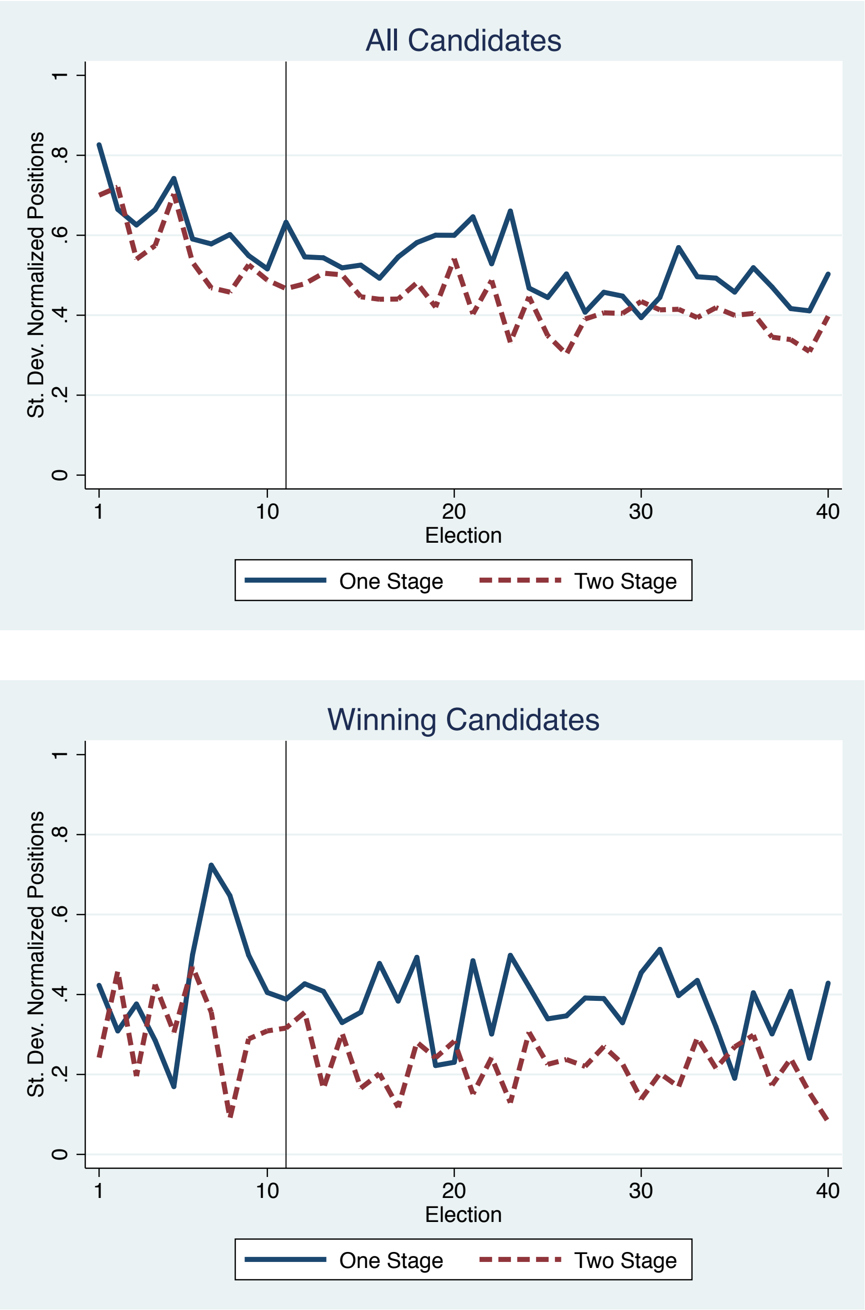


Figure 5: Disperson of positions and outcomes

candidates are close to the median voter (90%). Yet, there is an asymmetry because when the extremist candidate's position is extreme (closest to its own ideal point between 0.8 and 1 on the normalized scale), the choice between the moderate and extremist is essentially a coin flip (extremist wins 51% of the time). The more moderate candidates do best when they locate near the midpoint between their party's ideal point and the median voter and when the extremist is more extreme, but even then, the moderate does not do much better than a coin flip, winning elections at most 57% of the time (when the moderate is between 0.4 and 0.6 and the extremist is between 0.6 and 0.8).[19]

In Table 2, I characterize voting in each group of 10 elections according to three possible behavioral rules. The first row shows the percentage of votes for the moderate candidate. Notice that fewer 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 of primaries in elections 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.

However, primary voters do not express clear, unconditional preferences for either moderate or extremist party candidates. Table 2 therefore characterizes two additional behavioral rules. The second row shows that a simple "midpoint" strategy where voters select the candidate closest to the midpoint between the median voter and their party's position is a better description of behavior than voting for the moderate (or the extremist).

[19] See the Appendix for additional details about voting behavior as a function of the candidates' positions.

Table 2: Primary voting behavior

	Elections				
Voting rule	1-10	11-20	21-30	31-40	Total
Closer to median voter	37% (525)	44% (575)	46% (630)	46% (635)	43% (2,365)
Closer to midpoint	67% (645)	65% (640)	64% (680)	65% (655)	65% (2,620)
Closer to own promise	45% (648)	76% (644)	75% (671)	78% (652)	68% (2,615)

Roughly two-thirds of votes (overall 65%) are consistent with this rule (compared to the 43% consistent with voting for moderates and 57% voting for extremists).

The third row of Table 2 presents analysis of the third voting rule in which voters behave as if they have heterogeneous "belief-induced ideal points." This rule appears to be the most consistent with the data. It 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 the experiment, 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. This voting rule attains the highest rate of classification success, outperforming the simple midpoint rule in elections with feedback. 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 strategy. 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 (as discussed in the theoretical analysis and consistent with the patterns shown in Figure 5).

A Direct Test of Beliefs and Behavior

The experimental findings that candidate positions diverge from the median voter's position in both 1S and 2S elections supports the behavioral theory predicated on out-of-equilibrium beliefs (Prediction 3) over the competing predictions based on fully strategic candidate behavior (Predicitons 1 and 2). This inference is indirect, however, because beliefs are neither measured nor manipulated in the experiment. To generate a more direct test of the connection between beliefs and behavior, I conducted another version of the experiment in which beliefs are more carefully controlled and manipulated. In this version of the experiment, subjects play candidates in the 2S election game. Greater control over beliefs is achieved by having subjects play against computer opponents rather than other subjects. This ensures that the distribution of positions is known and exogenous. Variation in beliefs is induced by providing truthful information about whether the opposing party's candidate is moderate or extreme.

I conducted three sessions of the modified experiment (54 participants, 18 subjects per session). Each subject played 20 rounds of the 2S game (with feedback) against computer opponents.[20] The game is modified so that subjects know that the opposing candidate's position in the general (second stage) election is stochastically determined by a two-part process. First, two opposing primary candidates' positions are randomly drawn from a uniform distribution over the positions between the median voter's ideal point and the opposing party's ideal point. Second, the opposing party's primary voter (also the computer) randomly selects one of the two positions with equal probability. Information about whether the opposing computer voter chose the moderate or extremist is then provided to the subject. The con-

[20] The sessions were divided into three parts, with the candidate choices of interest coming in Part 3. The procedures for Part 1 are the same as in the main experiment for the 2S election game without feedback. This ensures that subjects have the same experience with the game as the subjects in the original experiment. The only minor differences are that groups have 9 players instead of 7 and the distance between parties is held constant at 120. Part 2 involves voting against random opponents, but those data are not analyzed here.

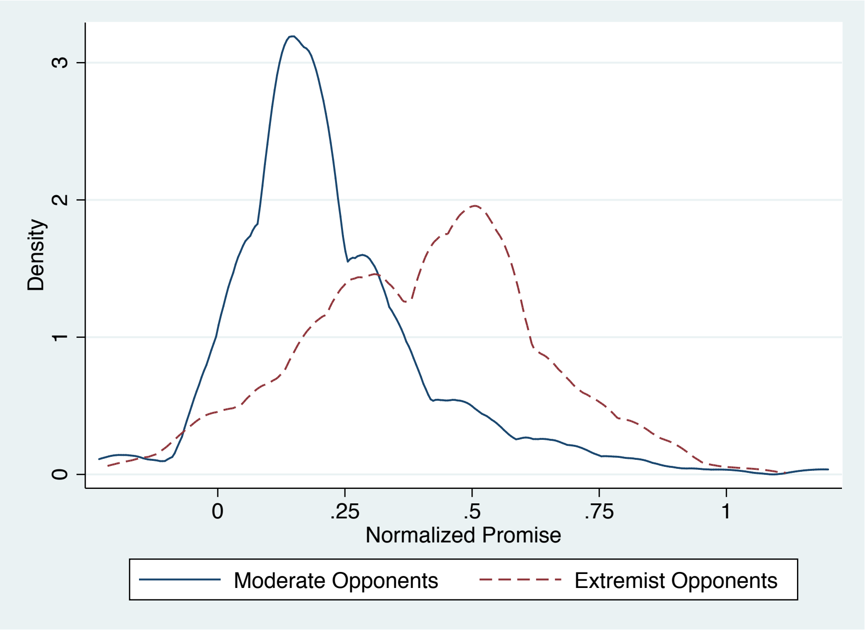


Figure 6: Candidate positions against random (computer-selected) moderates and extremists

sequence of this two-part procedure is that candidates will face one of the right-triangular distributions shown in the top-right and bottom-right of Figure 2. When the opponent is known to be more moderate, the distribution skews towards the median voter, and when the opponent is known to be more extreme, the distribution skews toward the opposing party's ideal point.

Figure 6 provides a comparison of the distributions, plotted as kernel densities, that candidates take against moderates versus extremists. The results demonstrate that subjects clearly respond to information about the extremity of their opponents. When computers select moderate opponents, human candidates take correspondingly moderate positions. Specifically, the majority of positions (69%) fall between 0 and 0.25 on the normalized scale (25% of the distance from the median to their own party's ideal point). When opponents are extreme, the distribution of positions shifts considerably towards their own party's ideal point (67% of the distribution shifts to *above* 0.25). In addition, the mode increases sharply, from within the interval between 0 and 0.25 to 0.5 on the normalized scale. The mean position shifts from 0.20 against moderates to 0.35 against extremists, and this difference is statistically significant ($p < 0.01$). These results provide direct evidence that candidates' positions respond to exogenously induced changes in their beliefs, supporting the belief-based behavioral theory.

Conclusion

Analytically, I strip away many of the complexities of real-world elections and focus on how introducing voters affects candidate positions in a stark environment with policy-motivated actors and complete information. When candidates and voters have imperfect, out-of-equilibrium beliefs about the behavior of their opponents, the extremity of the positions and candidates they choose depends on how extreme they expect their opponents to be. Thus, the effect of primary elections is conditioned by beliefs and strategic expectations.

Experimentally, I find that the need to win a partisan primary does not affect candidates' positions. To the extent there is any polarization, it occurs because primary voters select extremists more often than they select moderates—that is, through the behavior of voters rather than the strategic responses of candidates. The effect is relatively small and limited to settings in which participants cannot learn about the behavior of others from past experience.

Interestingly, the experimental data also demonstrate that primary elections may in fact contribute to "ideological purity," but not in the way that conventional wisdom suggests. Primary voters do not seem to care about ideological purity per se. Instead, 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. Candidates are responsive to this selective weeding out by voters and, as a consequence, take positions that are more homogeneous in elections with primaries than elections without them. Moreover, candidates are responsive to exogenous manipulations of opposing candidates' positions, which provides direct support for the mechanism posited by the theory.

The behavioral theory helps to make sense of the fact that many people blame partisan primary elections for much of the polarization and dysfunction that afflicts the contemporary American political system but empirical research has not been able to provide compelling evidence to support the claim. For example, the theory is consistent with the findings that neither the introduction of direct primaries (Hirano et al. 2010) nor the format of primary elections (McGhee et al. 2014) has much to do with increasing polarization. It is also consistent with the fact that polarization has been increasing over time despite the absence of significant changes in electoral institutions.

A direction for future observational research would be to explore the notion that strategic expectations about increasing polarization may, to some extent, be self-fulfilling. For example, the theory implies that partisans who increasingly perceive the opposing party's candidates to be more extreme will be emboldened to support more extreme candidates of their own (e.g., this may explain Bernie Sanders' popularity in the 2016 Democratic primary). With appropriate measures, this hypothesis could be tested both cross-sectionally or over time, and it is entirely plausible to the extent that citizens infer extreme ideological positions from their dislike of the opposing party (Brady and Sniderman 1985) given the steady rise of negative partisanship and affective polarization (Abramowitz and Webster 2016, Iyengar, Sood and Lelkes 2012).

References

Abramowitz, Alan I and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of US elections in the 21st century." *Electoral Studies* 41:12–22.

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* .

Aldrich, John H. and Skip Lupia. 2011. Experiments and Game Theory's Value to Political Science. In *Cambridge Handbook of Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge University Press.

Ansolabehere, Stephen, James M Snyder, Jr. and Charles Stewart, III. 2001. "Candidate positioning in US House elections." *American Journal of Political Science* 45(1):136–159.

Aragones, Enriqueta and Thomas R. Palfrey. 2007. "The Effect of Candidate Quality on Electoral Equilibrium: An Experimental Study." *American Political Science Review* 98(February):77–90.

Aronson, Peter H. and Peter C. Ordeshook. 1972. Spatial Strategies for Sequential Elections. In *Probability Models of Collective Decision Making*, ed. Richard G. Niemi and Herbert F. Weisberg. Charles E. Merrill pp. 298–331.

Aumann, Robert and Adam Brandenburger. 1995. "Epistemic conditions for Nash equilibrium." *Econometrica* 63(5):1161–1180.

Battaglini, Marco, Rebecca B Morton and Thomas R Palfrey. 2010. "The Swing Voter's Curse in the Laboratory." *The Review of Economic Studies* 77(1):61–89.

Belot, Michele, Raymond Duch and Luis Miller. 2015. "A comprehensive comparison of students and non-students in classic experimental games." *Journal of Economic Behavior & Organization* 113:26–33.

Bendor, Jonathan, Daniel Diermeier, David A. Siegel and Michael M. Ting. 2011. *A Behavioral Theory of Elections.* Princeton University Press.

Bonica, Adam. 2013. "Ideology and interests in the political marketplace." *American Journal of Political Science* 57(2):294–311.

Brady, David W., Hahrie Han and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1):79–105.

Brady, Henry E and Paul M Sniderman. 1985. "Attitude attribution: A group basis for political reasoning." *American Political Science Review* 79(4):1061–1078.

Bullock, Will and Josh Clinton. 2011. "More of a Molehill than a Mountain: The Effects of the Blanket Primary on Elected Officials' Behavior from California." *Journal of Politics* 73(3):915–30.

Callander, Steven and Catherine H. Wilson. 2007. "Turnout, Polarization, and Duverger's Law." *The Journal of Politics* 69(November):1047–1056.

Calvert, Randall. 1985. “Robustness of the Multidimensional Voting Model: Candidates’ Motivations, Uncertainty, and Convergence.” *American Political Science Review* 29(1):69–95.

Camerer, Colin. 2003. *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press.

Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong. 2004. “A Cognitive Hierarchy Model of Games.” *Quarterly Journal of Economics* 119(3):861–898.

Chen, Kong-Pin and Sheng-Zhang Yang. 2002. “Strategic voting in open primaries.” *Public Choice* 112(1-2):1–30.

Cherry, Todd L. and Stephan Kroll. 2003. “Crashing the Party: An experimental investigation of strategic voting in primary elections.” *Public Choice* 114:387–420.

Cho, Seok-Ju and Insun Kang. 2014. “Open primaries and crossover voting.” *Journal of Theoretical Politics* .

Coleman, James S. 1972. The Positions of Political Parties in Elections. In *Probability Models of Collective Decision Making*, ed. Richard G. Niemi and Herbert F. Weisberg. Charles E. Merrill pp. 332–357.

Collier, Kenneth E, Richard D McKelvey, Peter C Ordeshook and Kenneth C Williams. 1987. “Retrospective voting: An experimental study.” *Public Choice* 53(2):101–130.

Cooper, David J, John H Kagel, Wei Lo and Qing Liang Gu. 1999. “Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers.” *American Economic Review* 89(4):781–804.

Crawford, Vincent P. 2003. “Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intention.” *American Economic Review* 93(1):133–149.

Crawford, Vincent P, Miguel A Costa-Gomes and Nagore Iriberri. 2013. “Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications.” *Journal of Economic Literature* 51(1):5–62.

Dickson, Eric. 2011. Economics versus Psychology Experiments: Stylization, IIncentive, and Deception. In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge University Press.

Druckman, Jamie N. and Cindy D. Kam. 2011. Students as Experimental Participants: A Defense of the ”Narrow Data Base”. In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge University Press pp. 41–57.

Eckel, Catherine and Charles A Holt. 1989. “Strategic voting in agenda-controlled committee experiments.” *American Economic Review* 79(4):763–773.

Falk, Armin and James J Heckman. 2009. "Lab experiments are a major source of knowledge in the social sciences." *science* 326(5952):535–538.

Fatas, Enrique, Tibor Neugebauer and Pilar Tamborero. 2007. "How politicians make decisions: A political choice experiment." *Journal of Economics* 92(2):167–196.

Fiorina, M P and C R Plott. 1978. "Committee decisions under majority rule: An experimental study." *The American Political Science Review* .

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economics Experiments." *Experimental Economics* 10(2):171–8.

Frechette, Guillaume R, John H Kagel and Steven F Lehrer. 2003. "Bargaining in legislatures: An experimental investigation of open versus closed amendment rules." *American Political science review* 97(2):221–232.

Fudenberg, Drew and David K Levine. 1998. *The Theory of Learning in Games*. Vol. 2 MIT press.

Gerber, Elisabeth R. and Rebecca B. Morton. 1998. "Primary Election Systems and Representation." *Journal of Law, Economics, and Organization* 14(2):302–324.

Hacker, Jacob S and Paul Pierson. 2006. *Off center: The Republican revolution and the erosion of American democracy*. Yale University Press.

Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109(1):18–42.

Healy, Andrew and Neil Malhotra. 2009. "Myopic Voters and Natural Disaster Policy." *American Political Science Review* 103(August):387–406.

Herzberg, Roberta Q and Rick K Wilson. 1988. "Results on sophisticated voting in an experimental setting." *Journal of Politics* 50(2):471–486.

Hirano, Shigeo, James M. Snyder, Jr. and Michael M. Ting. 2009. "Distributive politics with primaries." *Journal of Politics* 71(4):1467–1480.

Hirano, Shigeo, James M. Snyder, Jr., Stephen Ansolabehere and John Mark Hansen. 2010. "Primary Elections and Partisan Polarization in the U.S. Congress." *Quarterly Journal of Political Science* 5(2):169–191.

Huber, Gregory A., Seth J. Hill and Gabriel S. Lenz. 2012. "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents." *American Political Science Review* 106(4):720–741.

Hummel, Patrick. 2013. "Candidate strategies in primaries and general elections with candidates of heterogeneous quality." *Games and Economic Behavior* 78:85–102.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, Not IdeologyA Social Identity Perspective on Polarization." *Public opinion quarterly* 76(3):405–431.

Jackson, Matthew O., Laurent Mathevet and Kyle Mattes. 2007. "Nomination processes and policy outcomes." *Quarterly Journal of Political Science* 2(1):67–92.

Mann, Thomas E and Norman J Ornstein. 2013. *It's even worse than it looks: How the American constitutional system collided with the new politics of extremism.* Basic Books.

McCarty, Nolan, Keith T Poole and Howard Rosenthal. 2006. *Polarized America: The dance of ideology and unequal riches.* MIT Press.

McCarty, Nolan, Keith T Poole and Howard Rosenthal. 2013. *Political Bubbles: Financial Crises and the Failure of American Democracy.* Princeton University Press.

McCuen, Brian and Rebecca B Morton. 2010. "Tactical coalition voting and information in the laboratory." *Electoral Studies* 29(3):316–328.

McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers and Nolan McCarty. 2014. "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology." *American Journal of Political Science* 558(2):337–351.

McKelvey, Richard D. and Peter C. Ordeshook. 1982. "Two-Candidate Elections without Majority Rule Equilibria An Experimental Study.".

McKelvey, Richard D and Peter C Ordeshook. 1985. "Sequential Elections with Limited Information." *American Journal of Political Science* 29(3):480–512.

Meirowitz, Adam. 2005. "Informational party primaries and strategic ambiguity." *Journal of Theoretical Politics* 17(1):107–136.

Mintz, Alex, Steven B Redd and Arnold Vedlitz. 2006. "Can we generalize from student experiments to the real world in political science, military affairs, and international relations?" *Journal of Conflict Resolution* 50(5):757–776.

Morton, Rebecca B. 1993. "Incomplete Information and Ideological Explanations of Platform Divergence." *American Political Science Review* 87(2):382–392.

Morton, Rebecca B. and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab.* Cambridge University Press.

Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *The American Economic Review* 85(5):1313–1326.

Oak, Mandar P. 2006. "On the role of the Primary System in Candidate Selection." *Economics & Politics* 18(2):169–190.

Ostrom, Elinor. 1998. "A Behavioral Approach to the Rational Choice Theory of Coallective Action: Presidential Address, American Political Science Association, 1997." *American Political Science Review* 92(1):1–22.

Owen, Guillermo and Bernard Grofman. 2006. "Two-stage electoral competition in two-party contests: persistent divergence of party positions." *Social Choice and Welfare* 26(3):547–569.

Palfrey, Thomas R. 2006. Laboratory Experiments. In *Oxford Handbook of Political Economy*, ed. Barry R. Weingast and Donald A. Wittman. Oxford University Press pp. 915–936.

Plott, Charles R and Michael E Levine. 1978. "A model of agenda influence on committee decisions." *American Economic Review* 68(1):146–160.

Poole, Keith T and Howard Rosenthal. 1984. "The polarization of American politics." *The Journal of Politics* 46(4):1061–1079.

Poole, Keith T and Howard Rosenthal. 1997. *Congress: A political-economic history of roll call voting*. Oxford University Press.

Potters, Jan and Frans van Winden. 2000. "Professionals and Students in a Lobbying Experiment: Professional Rules of Conduct and Subject Surrogacy." *Journal of Economic Behavior & Organization* 43:499–522.

Rietz, Thomas. 2008. "Three-way experimental election results: strategic voting, coordinated outcomes and Duverger's law." *Handbook of Experimental Economics Results* 1:889–897.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69(1):99–118.

Smirnov, Oleg. 2009. "Endogenous choice of amendment agendas: types of voters and experimental evidence." *Public Choice* 141(3-4):277–290.

Snyder, Jr., James M. and Michael M. Ting. 2011. "Electoral selection with parties and primaries." *American Journal of Political Science* 55(4):782–796.

Stahl, Dale O. and Paul W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior* 10(1):218–254.

Van der Straeten, Karine, Jean-Francois Laslier, Nicolas Sauger and Andre Blais. 2010. "Strategic, Sincere, and Heuristic Voting under four Election Rules: An Experimental Study." *Social Choice and Welfare* 35(3):435–472.

Wittman, Donald A. 1983. "Candidate Motivations: A Synthesis of Alternative Theories." *American Political Science Review* 77(1):142–57.

Woon, Jonathan. 2012. "Democratic Accountability and Retrospective Voting: A Laboratory Experiment." *American Journal of Political Science* 56(4):913–930.

Online Appendix

Table A3 provides a more detailed description of voting behavior as a function of the candidates' positions than is described in the main text of the paper. 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.

Table A3: Votes for moderate by candidates' positions

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

Fixed Roles Experiment

Procedures

I designed and conducted a different version of the experiment in an effort 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 primarily by assigning subjects to fixed roles. Instead of choosing positions in each round before candidates are selected (as the main experiment), subjects are randomly assigned to roles as candidates and voters *before* the first election and then retain their roles throughout the experiment. In the 1S condition of the fixed role experiment, all subjects are candidates and are randomly matched in pairs for each election (one left candidate against one right candidate, with no subjects as voters). In the 2S condition, 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 fixed role experiment is a close analogue to Part 2 of main experiment (the 30 elections with feedback). I conducted two sessions of the fixed role experiment with 1S elections (36 subjects) and three sessions with 2S elections (48 subjects) at the *******.

Elections 31-50 of the fixed role experiment 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, 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. However, this setup does not manipulate beliefs or information directly the way that the direct test does in the main text of the paper. 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 with candidates; 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 A1 shows the average positions over time in the fixed roles experiment, plotted separately for 1S and 2S elections. In contrast to the original setup, primaries with fixed roles cause candidates to take more extreme positions than they do in 1S elections. The top panel of Figure A1 suggests that this effect is modest but persistent over time. Similar to the original experiment, 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

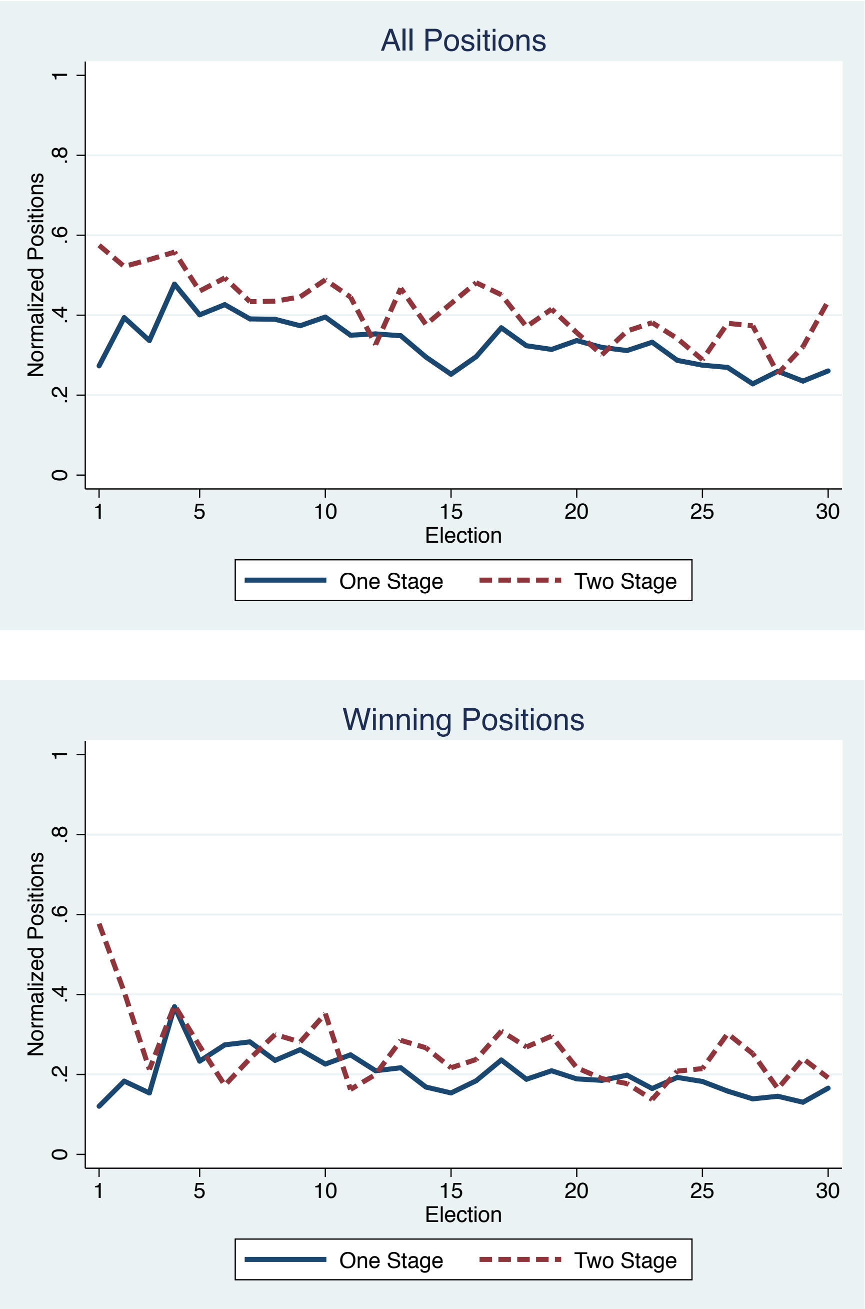
Table A4: Regression analysis of positions in the Fixed Roles Experiment

	(1) All	(2) Party	(3) Winner
Primary (2S) Elections	0.088**	0.056**	0.057**
	(0.015)	(0.017)	(0.016)
Experience	-0.006**	-0.006**	-0.004**
	(0.001)	(0.001)	(0.001)
Constant	0.417**	0.415**	0.253**
	(0.016)	(0.016)	(0.015)
Observations	2040	1560	780
R^2	0.0391	0.0338	0.0400

* < .05 ** < .01

the last five rounds, the average in 1S elections diminishes to .251 compare to .353 in 2S elections. The regression analyis in Table A4 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 signficant 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 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 the fixed roles experiment than it was in the original (64% of votes are for moderates against human players and 66% are for moderates against random positions, $p = .59$). These results suggest that candidates in the fixed roles experiment are primarily level-1, choosing positions



All Positions

Winning Positions

Figure A1: Average positions and outcomes in the Fixed Roles Experiment

as if their opponents choose their positions randomly (level-0), providing some additional support for the behavioral theory.

Online Appendix: Instructions for Two-Stage Elections (with Fixed Matching)

Instructions

General Information

This is an experiment on the economics of elections. The ******* 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.

Online 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).

Online 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:

Points = 200 - |Winning campaign promise – 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.

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:

   Points = 200 - |Winning campaign promise – 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?

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.

Online Appendix: Instructions for Direct Test of Belief Manipulation

Instructions

General Information

This is an experiment on the economics of elections. The ********** has provided the 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. More specifically, we will take the total number of points you earned in the election that counts, divide by 10, and then round this amount to the nearest quarter. You should think of each election as a separate decision task.

Online Appendix: Instructions for Direct Test of Belief Manipulation

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 your position is selected as the winner of the election, then this number will determine the payoffs for each participant (as we will explain later).

Once all of the participants have chosen their campaign promises, the computer will then randomly select two promises from each group (with each member's promise equally likely to be selected), and then we will move to the voting stages. Note that even though only two promises from your group will be selected, you should choose your campaign promise as if it were actually selected.

First Voting Stage

In the first voting stage, each group votes to select one of the group's two promises to put forward for the second voting stage. The promise that receives the most votes will move on to the second voting stage. If the vote is a tie, each promise is equally likely to be selected. Only members of your group will be voting on the promises from your own group in the first voting stage. The other group will be voting on their own two promises at the same time.

Second Voting Stage

In the second voting stage, a "computer voter" chooses the winning promise from the promises put forward by the groups. 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 promises in the second stage offer the computer voter the same payoff, then the computer voter will cast its vote randomly between the two promises (with each promise being equally likely to win).

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:

Points = 200 - |Winning campaign promise – Your favorite position|

For example, if your favorite position is 57 and the campaign promise selected by the computer voter in the second stage 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 players (including the computer voter) all earn points according to the same formula. That is, only the numerical value of the promise matters—you do not earn extra points if your promise is selected. 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 stage, and then you will cast your votes in all of the first voting stages. Note that you will not receive any feedback about the 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 Part 1 of the experiment.

2. Before every election, you will find out the favorite position of your group, the other group, and the computer voter.

3. In the campaign stage, you choose a number from 1 to 200 that serves as your campaign promise. If your promise wins the election, this campaign promise determines everyone's payoff.

4. We randomly select two promises from each group to be the choices in the first voting stage.

5. In the first voting stage, each group simultaneously votes for one promise to put forward in the second voting stage.

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

7. Your payoff is:

   Points = 200 - |Winning campaign promise – 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 100 and the winning campaign promise is 140, how many points would you earn?

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

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

4. Suppose that the computer voter's favorite position is 75. If Candidate A's campaign promise is 50 and Candidate B's campaign promise is 125, which candidate would the computer vote for?

5. If the computer voter's favorite position is 120, Candidate C's campaign promise is 60 and Candidate D's campaign promise is 150 which candidate would the computer vote for?

6. Suppose that the computer voter's favorite position is 80 and your favorite position is 20. If the computer voter's choice in the second stage is between Candidate E's campaign promise of 50 and Candidate F's campaign promise of 120, how many points will you earn?

Online Appendix: Instructions for Direct Test of Belief Manipulation

Part 2

There are 20 elections in Part 2. The basic rules for elections in Part 2 are similar to Part 1, and the payoffs will be determined in the same way as before. However, there are a few differences.

In every election, you will be paired against one voter from the other group, and your task is to vote in the first stage election. Each pair of voters interacts separately. That is, your payoffs will depend only on your actions and the actions of the voter you are paired against. The choices of other pairs of voters will not affect your payoffs in Part 2.

Instead of choosing your own campaign promises, you will choose between two promises that the computer randomly selects between your favorite position and the computer voter's favorite position. Each position in this range is equally likely to be selected as a campaign promise. For example, if your favorite position is 20 and the computer voter's favorite position is 80, then the computer will randomly select two promises between 20 and 80 and you will vote for one of these two promises.

Likewise, the other voter will choose between two promises randomly selected between their favorite position and the computer voter's favorite position. For example, if the computer's favorite position is 80 and the other voter's favorite position is 140, then the computer will randomly select two promises between 80 and 140 and the other voter will vote for one of these two promises.

As before, the computer voter will vote for the campaign promise closest to its own favorite position, and the campaign promise selected in the second voting stage determines everyone's payoff for that election.

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

Instruction Questions

1. How are the campaign promises chosen in Part 2?

2. If your favorite position is 40 and the computer voter's favorite position is 100, what is the range of values for the two randomly selected promises YOU will choose from?

3. If the other voter's favorite position is 150 and the computer voter's favorite position is 90, what is the range of values for the two randomly selected promises the OTHER voter will choose from?

4. Which rule does the computer voter use to determine the winner of the election?

Online Appendix: Instructions for Direct Test of Belief Manipulation

Part 3

There are 20 elections in Part 3. Instead of voting in the first stage election, your task is to choose a campaign promise from 1 to 200 (in the same way as in Part 1). Instead of playing against another participant, you will play against two computer players. One computer player will choose the other group's position, and a separate computer player will choose the winning position in the second voting stage just like in Parts 1 and 2.

As in Part 2, the other group will choose between two randomly selected campaign promises. Before choosing your promise, you will be given a piece of information about which position the other group's voter selected. Specifically, you will find out if the voter selected the promise that is closer to the computer voter's favorite position or the promise that is closer to the other group's favorite position. However, you will not know the exact numerical value of the position.

For example, suppose the other group's favorite position is 30 and the computer voter's favorite position is 90 while the other group's two promises are 40 and 80. There are two possible pieces of information you might receive:

- If the other group chooses 40, then you will see a message saying "the other voter chose the promise closer to the OTHER voter's favorite position" (since 40 is closer to 30 than 80 is).

- On the other hand, if the other group chooses 80, then you will see a message saying "the other voter chose the promise closer to the COMPUTER voter's favorite position" (since 80 is closer to 90 than 30 is).

Otherwise, the rules for determining your payoffs (such as the payoff formula) and for determining the computer voter's vote are the same as in Parts 1 and 2.

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

Instruction Questions

1. How is the other group's campaign promise determined in Part 3?

2. Suppose the other group's promises were 130 and 145, and that you also know that the (second stage) computer voter's favorite position is 100 and the other group's favorite position is 160. Which of these promises is closer to the (second stage) COMPUTER voter's favorite position?

3. Suppose the other group's promises were 30 and 60, and that you also know that the (second stage) computer voter's favorite position is 80 and the other group's favorite position is 20. Which of these promises is closer to the OTHER group's favorite position?