How to Elect More Women:
Gender and Candidate Success in a Field Experiment

Christopher Karpowitz, Quin Monson, and Jessica Preece[1]
Department of Political Science
Brigham Young University

Abstract
Women are dramatically underrepresented in legislative bodies, and most scholars agree that the greatest limiting factor is the lack of female candidates (supply). However, voters' subconscious biases (demand) may also play a role, particularly among conservatives. We designed an original field experiment to test whether it is possible to increase women's electoral success through political party leaders' efforts to exogenously shock the supply of female candidates and/or voter demand for female representatives. The key experimental treatments involved messages from a state Republican Party chair to the leaders of 1,842 precinct-level caucus meetings. We find that party leaders' efforts to stoke both supply and demand (and especially both together) affect the number of women elected as delegates to the statewide nominating convention. We then replicate this finding with a national sample of validated Republican primary election voters (N=2,897) using a vignette survey experiment. Our results suggest that simple interventions from party leaders can affect the behavior of candidates and voters and ultimately lead to a substantial increase in women's electoral success.

Word count: 9,828

---

[1] Corresponding author: jessica_preece@byu.edu

Introduction

Women are dramatically underrepresented in representative bodies, and research suggests that this may have important substantive implications for both policy and process (Karpowitz and Mendelberg 2014; Wängnerud 2009). In the United States, women comprise about 20 percent of Congress, 24 percent of state legislators, 12 percent of governors, and 17 percent of mayors of large cities.[2] Women's underrepresentation is particularly acute in the Republican Party, with the percentage of Republican state legislators in 2013 who are women stagnating at only 16.5 percent—a rate nearly identical to that of the early 1980s (Carroll and Sanbonmatsu 2013, 68).[3]

As norms about equality have shifted, political party elites have become increasingly concerned about the lack of female politicians in their ranks and have sought ways to rectify this imbalance.[4] Internationally, many parties have addressed this by implementing quotas for women (Krook 2009), but these are impractical in some electoral systems and anathema to parties with more conservative ideologies. Hence, it remains an open question whether parties facing these constraints have the ability to take action to increase women's representation in their parties.

We argue that party leaders, even party leaders of conservative parties, have effective levers for increasing the number of women in their ranks should they choose to use them. Specifically, party elites can focus on 1) increasing the *supply* of female candidates through active recruitment and 2) stoking *demand* for female representatives by setting a credible norm

---

[2] Center for American Women and Politics, "Current Numbers," http://www.cawp.rutgers.edu/current-numbers, accessed October 23, 2015.

[3] The proportion of state legislators who are Democrats as of 2013 is 32.3 percent, and has increased steadily, while the proportion among Republicans has not increased much since 1979.

[4] See, for example, the Republican National Committee's Growth and Opportunity Project report released after the 2012 presidential election: http://goproject.gop.com/.

of equality. Especially when done together, these elite behaviors have the potential to substantially increase women's representation.

To test this argument, we ran a field experiment with the cooperation of a state Republican Party in a state with low levels of women's representation. Party leaders were concerned that although about half of the party activists who attended neighborhood caucus meetings were women, generally only 20-25 percent of the delegates elected from these meetings to attend the state nominating convention were women. A field experiment conducted through party leaders in the context of a real election offers distinct advantages: high levels of realism, meaningful behavioral outcome measures, and solid causal inference. We randomly assigned over 2,000 precinct chairs to receive one of four letters from the state party chair prior to these neighborhood caucus meetings. The treatments were a neutral placebo control (Control), a request to recruit 2-3 women to run as state delegates (Supply), a request to read a letter at the precinct meeting encouraging attendees to elect more women as delegates (Demand), and a request to both recruit women and read the letter (Supply+Demand).

We find that messages from party elites had a potent effect on electoral outcomes. While both the Supply and Demand treatments somewhat increased the number of women elected as delegates, the Supply+Demand treatment had the largest and most statistically powerful effect on the number of women elected. The percentage of precincts that elected at least one female delegate jumped from 37.5 percent in the Control condition to 45.4 percent in the Supply+Demand condition. Further, in precincts assigned to control conditions, women comprised about 24 percent of delegates, while under the Supply+Demand condition women comprised close to 30 percent of delegates. These findings are robust to a wide variety of model specifications and suggest that simple encouragement interventions from party leaders may have

the ability to increase the number of women in office. To our knowledge, this is the first field experiment to demonstrate the ability of messages from party leaders to affect gender equality in a meaningful election setting.

We also ran a vignette experiment on a large national sample of validated Republican primary election voters (N=2,897). As in the field experiment, some respondents were presented with statements from Republican Party officials encouraging them to elect more women (Demand), some saw a slate of candidates with additional female candidates to choose from (Supply), and some saw both (Supply+Demand). Confirming that our field experiment results are not simply context specific, we find a large and statistically significant effect on respondent vote choice for female candidates in the Supply+Demand condition compared to the control. Active party recruitment of female candidates combined with credible party rhetoric encouraging voters to vote for women can have an important impact on women's representation.

Literature Review and Expectations

Understanding why there are so few women in office requires thinking about both the *supply* of and *demand* for women on the ballot. While each of these likely influences women's representation, it is also possible that they work together in interesting ways.

The Supply of Female Candidates: Political Ambition and Recruitment

The first factor that influences the supply of female candidates on the ballot is how many women *want* to run for office. Most studies find significantly lower levels of political ambition among women than among men. Explanations include women's family obligations (Fulton et al. 2006; Sapiro 1982), gender role socialization (Clark, Hadley, and Darcy 1989), differences in qualifications and perceptions of qualifications (Crowder-Meyer and Lauderdale 2014; Fox and Lawless 2004, 2005; Lawless and Fox 2010), differences in party support and perceptions of

party support (Fox and Lawless 2004, 2010, 2011; Fulton et al. 2006; Lawless and Fox 2010; Sanbonmatsu 2006a, 2006b), and competition aversion (Kanthak and Woon 2013; Preece and Stoddard 2015).

Women's lower levels of political ambition may actually be endogenous to gendered patterns in political party recruitment. Recruitment is an important aspect of the candidate emergence process in American politics (Broockman 2014; Fox and Lawless 2010; Maestas et al. 2006); however, women are much less likely to be recruited than men (Crowder-Meyer 2013; Fox and Lawless 2010; Lawless and Fox 2010; Niven 1998; Sanbonmatsu 2006a, 2006b). The recruitment gender gap is particularly important because women tend to make decisions about running for office through a "relationally embedded" process where decisions are sensitive to social cues and relationships, rather than one based on pure political ambition (Carroll and Sanbonmatsu 2013).

Hence, increased political party efforts to recruit women are likely to play an important role in increasing women's representation. Among women and men who report being recruited by political elites, Lawless and Fox (2010) find no difference in levels of political ambition. All of this leads us to expect that an experimental intervention focused on recruiting women to run to for office should increase the number of women elected to these positions.

Demand for Female Candidates: Voter Stereotypes and Vote Choice

Voters' willingness to support female candidates has significantly increased over the last few decades, though Republicans are significantly less likely to favor gender parity in government than Democrats (Dolan and Sanbonmatsu 2008), and implicit gender attitudes toward female leadership remain (Mo 2014). For example, because voters tend to believe that male-stereotyped characteristics are more important than female-stereotyped characteristics in

politics, these stereotypes often hurt voters' perceptions of female candidates (Brown, Heighberger, and Shocket 1993; Fox and Smith 1998; Huddy and Terkildsen 1993; Lawless 2004). Likewise, policy stereotypes, such as whether men or women are better at handling the economy or national security, may harm women (Dolan 2010; Lawless 2004).[5]

Given these stereotypes, Republican female candidates face a particularly interesting set of electoral challenges (Dolan and Lynch 2013). Because they are perceived to be more liberal than male Republican candidates, they may be more appealing to the median general election voter (King and Matland 2003). However, this stereotype actually hurts them among fellow Republicans (King and Matland 2003; Koch 2002; Sanbonmatsu and Dolan 2009). This issue of "party fit" plays a significant role in explaining why there are so few Republican women in office (Thomsen 2014, 2015).

Is there anything the party can do about this? Little research exits about whether parties have the ability to influence voters' demand for female candidates. Yet, we know that partisans—especially well-informed partisans—pay close attention to ideological cues from party elites (Zaller 1992). Party elites are "knowledgeable and trustworthy" sources who are likely to be persuasive among their followers (Lupia and McCubbins 1998, 201). Because of this dynamic, Republican elites' decisions to distance the party from the women's rights movement may have created a party culture that helps to explain why Republican women's representation actually peaked in the late 1990s (Carroll and Sanbonmatsu 2013; Elder 2012, 2014).

---

[5] Stereotypes can also benefit women, such as in some kinds of negative campaign environments (Fridkin, Kenney, and Woodall 2009; Krupnikov and Bauer 2014) and among the subset of voters who have a preference for female candidates (Dolan 2010; Sanbonmatsu 2002). Some trait-based stereotypes, such as whether men or women are more assertive or compassionate, are unlikely to disproportionately harm women on Election Day (Brooks 2013; Dolan 2010).

Hence, if Republican elites sent a clear message to partisans that women's representation was a priority for the party, perhaps they would follow their lead.[6] The mechanism for the relationship between messages of party leaders and caucus voting behavior is norms or expectations about the role of women in politics. Politics is still seen by many as a "man's game" (Burns, Schlozman, and Verba 2001), and this gendered expectation may be especially strong among Republicans, given the party's history of ambivalence toward the women's movement. When trusted Republican Party leaders send a clear, persuasive message that women's voices are welcome and needed, we expect that partisans will follow the cue that voting for women is consistent with the party's values. Such responsiveness to the messages of party leaders may occur because the party leaders are reminding their fellow partisans of a norm the voters already hold (gender equality) and emphasizing that the norm applies to the intra-party context or because party leaders are effective in establishing new expectations for the group (in contrast to earlier ambivalence about women's representation). Regardless, credible statements from party leaders in support of women's representation sets a norm that is likely to get a response from voters.

Field Experiment Research Design

We worked with the state Republican Party in a Republican-dominated state with low levels of women's representation to see how party attempts to exogenously shock the supply of and demand for female candidates affect women's political participation. The state is one of 15 in which women comprise less than 20 percent of the state legislature and one of 18 in which women comprise less than 15 percent of Republican state legislators.

---

[6] In fact, the Republican National Committee's post 2012 election Growth and Opportunity Project report did identify increasing the number of female Republican politicians as a priority. See http://goproject.gop.com/.

The experimental setting was the party's caucus meetings, in which neighborhood precincts elect delegates to the state party nominating convention. Despite the fact that men and women attend the caucus meetings in approximately equal numbers, women have been dramatically under-represented at the party convention, generally comprising only about 20-25 percent of the elected delegates. This imbalance is especially striking because caucus-goers regularly elect women to less prestigious positions in high numbers, including secretary, treasurer, and delegates to the county party conventions. Hence, our interventions were designed to increase the number of women elected to be delegates to the state nominating convention, the most important and contested position elected at the caucus meeting.

Prior to the caucus meetings in March 2014, we received a list of the precinct chairs from the state Republican Party. Precinct chairs lead the neighborhood caucus meetings, so the list allowed us to communicate directly with the person responsible for organizing and conducting each meeting. The experiment is a 2x2 factorial design in which party precinct chairs were randomly assigned to receive one of four letters printed on the party's letterhead and signed by the State Party Chair. The full text of the letters for each condition is available in Supporting Information Appendix B.

- *Placebo Control:* Encouraged the precinct chair to foster good environment at the caucus so individuals would feel comfortable running for positions. This condition included no mention of gender at all.
- *Supply:* After an introduction that reviewed the gender disparity among state delegates, the precinct chair was encouraged "to reach out and encourage two or three women in your precinct to run for positions as precinct leaders or delegates this year."
- *Demand:* After an introduction that reviewed the gender disparity among state delegates, the precinct chair was encouraged to read a paragraph at beginning of caucus meeting that reviewed the gender disparity and concluded with a plea to "consider this as you nominate and vote for delegates. We would like to see more women at our state convention."
- *Supply+Demand:* After an introduction that reviewed the gender disparity among state delegates, the precinct chair was encouraged to reach out to two or three women ***and*** read the paragraph.

Our experiment falls into the category of an encouragement design, common in public health and get-out-the-vote experiments (Gerber and Green 2000). Encouragement designs are distinctive because they focus on the effect of asking subjects to participate in a treatment, rather than the effect of the treatment itself. While this approach can lead to challenges with compliance, encouragement designs reflect the reality of many kinds of real world interventions (Gerber and Green 2012). Because of the challenges that typically come from two-sided non-compliance in encouragement designs, we have chosen to focus our analysis on the conservative "intent-to-treat" analysis (Gerber and Green 2012, 173-205).[7] Furthermore, our intervention is encouraging precinct chairs to perform tasks, while our outcome measures depend on the behavior of all precinct caucus attendees, most of whom were not directly treated. This situates the study between traditional experimental studies of mass behavior and newer experimental studies of elites and institutions (Grose 2014; Loewen, Rubenson, and Wantchekon 2010). Hence, this experiment is best thought of as an examination of the downstream mass partisan effects of encouraging party elites to take action (Gerber and Green 2012, 193-205).

Mailed letters were the most inclusive, simple, and consistent way to contact all precinct chairs. Unfortunately, when we received the list of precinct chairs approximately two weeks before caucus night, the party's list of mailing addresses was still incomplete, and our mail services contractor determined that another handful of addresses was outdated (the addressee had moved, for example) or otherwise incorrect. Of the 2,156 precincts in the state, we were able to send letters to 1,842 precincts. Table 1 shows how many precincts were assigned to each condition and the proportion with sufficient information to send the treatment letters. As the

---

[7] This choice is also appropriate because we do not have compliance data for all caucus precincts in the state. As we explain below, we have compliance information for only a subset of the precincts, making computing the statewide complier average causal effect (CACE) difficult.

table reveals, precincts with faulty or no mailing addresses were equally distributed across the experimental conditions.

Table 1: Assignment to Conditions

Condition	# in Condition	# Sent Letter	% Sent Letter
Control	541	465	85.95%
Supply	539	456	84.60%
Demand	538	462	85.87%
Supply + Demand	538	459	85.32%
Total	2,156	1,842	85.44%

Randomization

To evaluate the success of our random assignment procedures, we compared the characteristics of the precincts across the conditions. We conducted four separate tests: a Wilcoxon-Mann-Withey test, a two-sample difference-of-means test, a Kolmogorov-Smirnov test, and an equality of medians test.[8] Table 2 shows how a variety of precinct-level characteristics compare across the conditions. Across all four statistical tests, only one comparison reaches standard levels of statistical significance: the difference in average distance from the convention site between the control and the Demand condition in a two-sample t-test. None of the other tests reveal a statistically significant difference between the conditions for that variable or any other. Given the large number of tests conducted, we would have expected some statistically significant results merely by chance, so we take the overwhelming number of null results as strong evidence that the randomization worked well.[9] Moreover, whatever small

---

[8] The first three tests include pairwise comparisons of every combination of experimental conditions, and the equality of medians test produces a chi-square statistic for all conditions simultaneously, comparing the number of precincts above and below the median in each condition.

[9] As an added test, we also regressed the precinct's 2012 proportion of state delegates who were women on the experimental conditions and controls for precinct-level characteristics. We again find no significant results (see Supporting Information Table A1). Two years prior to the experimental intervention, there was no relationship between women's representation and the experimental treatments.

differences we do find work *against* the prospects of finding results from the experimental treatments. For example, the fact that precincts in the control condition are slightly closer to the convention site should enable women's participation in control precincts, not deter it.

Table 2: Randomization Checks

	Placebo Control	Supply Condition	Demand Condition	Supply + Demand Condition
Percent Female State Delegates in 2012	24.8	23.2	24.1	24.5
Percent Precincts with Female Chair	25.1	24.7	21.8	25.7
Number of Delegate Positions Available	1.8	1.9	1.9	1.8
Mean Age of Caucus Participants	54.1	54.6	53.9	54.0
Number of Caucus Participants	23.4	24.3	24.6	23.0
Percent Caucus Participants Who Are Women	45.2	44.7	44.4	43.7
Distance between Precinct and Convention Site	55.0^c	63.7	65.5^a	62.6

Note: Cell entries are mean levels of each precinct-level variable, by experimental condition. Superscript letters indicate a statistically significant difference between conditions: *a* = difference vs. Control, *b* = difference vs. Supply, *c* = difference vs. Demand, *d* = difference vs. Supply + Demand.

Compliance

To understand the extent to which precinct chairs followed through on the requests by the state party chair, we designed a post-caucus survey of precinct chairs.[10] Between email and mail invitations, 1,901 precinct chairs were sent at least one invitation, 613 accessed the survey, and 604 precinct chairs completed the full survey for a 32 percent response rate.[11] Because we could not directly attend and observe all precinct meetings, let alone observe the efforts of precinct chairs to recruit female candidates prior to the meeting, the self-reports of the precinct chairs are our best available measure of compliance.

Table 3 highlights self-reported compliance among the 400 survey respondents who remembered receiving a letter from the precinct chair prior to caucus night as well as the same

---

[10] See Supporting Information Appendix C for additional details.

[11] Our response rate limits our compliance measurement strategy. Survey respondents may display selection bias because they are more conscientious in complying with the treatments (just as they complied with a request to complete a survey) or subject to social desirability pressures to report having carried out the treatments.

compliance results for all respondents to the survey.[12] In each condition, the patterns are much

as we anticipated. For each action – recruiting women or reading the paragraph encouraging

caucus attenders to elect women or both – compliance levels are highest in the experimental

conditions that correspond to those actions.

Table 3: Compliance Summary Statistics – Survey of Precinct Chairs

Condition	Obs	Reported Supply	Reported Demand	Reported Supply + Demand
Respondents Who Reported Receiving Letter Prior to Caucus Night (n=400)				
Control	93	11.8%	3.2%	2.2%
Supply	87	51.7%	19.5%	14.9%
Demand	114	43.9%	50.0%	24.6%
Supply + Demand	106	61.3%	41.5%	31.1%
All Respondents (n=604)				
Control	147	7.5%	2.0%	1.3%
Supply	155	31.0%	11.6%	9.0%
Demand	149	33.6%	38.3%	18.8%
Supply + Demand	153	43.1%	29.4%	21.6%

Source: 2014 Survey of Precinct Chairs

The results also show some spillover across conditions, as is often the case in

encouragement designs. For example, some level of recruitment (Supply) occurred entirely

outside of our efforts, with nearly 12 percent of respondents in the Control condition reporting

that they encouraged women to run as candidates for state delegates. This is not surprising,

however, because precinct chairs would reasonably consider it part of their duty to encourage

qualified candidates to stand for election as state delegates. However, nearly 44 percent of

respondents in the Demand condition told us that they also recruited female candidates,

significantly more than in the Control condition. It is reasonable that those planning to read the

Demand paragraph felt responsible to make extra efforts to ensure that female candidates would

---

[12] Some respondents could not remember whether or not they received a letter or reported receiving the letter after caucus night occurred.

be present and ready to run. Alternatively, perhaps those who received the letter with the encouragement to read the Demand paragraph felt increased need to report recruitment efforts, even if they did not actually engage in such efforts.

It is difficult to determine whether the spillover is the result of social desirability pressures or actual spillover in which precinct chairs discussed the letters' contents with each other, but the timing and mode of delivery minimized the chances of actual spillover.[13] Neither we nor the party received any communication that indicated that chairs had seen conflicting letters. Hence, we believe that most of the reported spillover is either because of social desirability or enterprising interpretations of the instructions. Assuming the reported two-sided non-compliance is genuine, it would also muddy the distinction between our treatments, and therefore make it more difficult to see treatment effects. This gives us confidence that we can trust the true treatment effects to be at least as large as we estimate. Importantly, our replication survey experiment below has no spillover potential, giving us a clean estimate of very similar treatment effects and casting additional doubt on the possibility that our field results are an artefact of two-sided non-compliance.

Field Experiment Results

We focus on two dependent variables: 1) the proportion of precincts in each condition that elected at least one woman to be a state delegate and 2) the proportion of state delegates in each condition who are women. The first is important because the majority of precincts in this sample do not elect any women; the second is a more traditional measure of women's

---

[13] The choice of postal mail delivery rather than email as the mode for the experimental treatment had the added benefit of minimizing the chance that precinct chairs would forward their letter to other chairs. We also waited to mail the letters until after the party's webinar training sessions to prevent discussion of them in these meetings. Most chairs received their letters 3-5 days before the caucus meeting, limiting the amount of time they had to share it with others. Our efforts to reduce spillover may thus have limited opportunity for compliance as well.

representation. Figure 1 shows the distribution of caucus attendees who are women compared to

the proportion of elected delegates who are women for the control condition. While attendees

are somewhat normally distributed between men and women, the modal outcome is electing zero

female delegates.

Figure 1: Distributions of Proportion of Women Who Are Caucus Attenders and Proportion of Women Elected as State Delegates
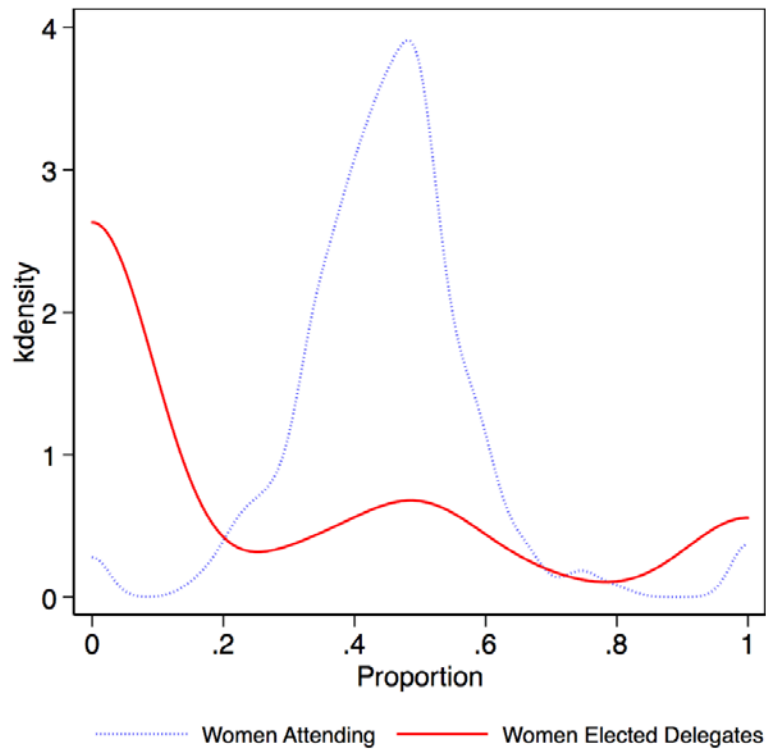
Note: Distributions for Control Condition Only.

Proportion of Precincts Electing At Least One Woman

Because most precincts elected more than one state delegate and the modal precinct

elected no women at all, we begin with the question of whether any of the elected delegates were

women.[14] Table 4 presents the percentage of precincts in each that elected at least one woman,

---

[14] The state delegate allocation for each precinct is determined by each county party and is a function of the relative Republican strength in the precinct as indicated by the number of Republican votes cast for statewide offices (Governor, Treasurer, Auditor, and Attorney General) in the most recent election.

computed from an analysis of the main experimental factors (Supporting Information Table A1).[15] Because of the full factorial research design, we can compare conditions that included each of the main factors with those conditions that were not exposed to those factors. For example, 40.4 percent of precincts that did not receive the Supply treatment elected at least one woman as a state delegate; that increased to 44.4 percent of precincts assigned to the Supply treatment, a statistically significant difference. The comparable numbers for those that were not assigned to read the Demand appeal (Control and Supply only) versus those that were (the Demand and Supply+Demand conditions) are 40.5 percent and 44.3 percent. In short, Table 4 offers clear evidence that the experiment's main treatment arms increased the proportion of precincts electing at least one woman.

Table 4: Main Effects of Treatments: Percent of Precincts Electing At Least One Woman

Supply		Demand	
Supply Treatment	44.37	Demand Treatment	44.33
N=913	(4.47)	N=889	(4.76)
No Supply Treatment	40.35	No Demand Treatment	40.49
N=899	(3.89)	N=923	(3.64)
Effect of Supply Treatment	4.02	Effect of Demand Treatment	3.84
	(1.56)		(1.87)
	[0.01]		[0.04]
95% Confidence Interval	0.97 to 7.08		0.18 to 7.50

Note: Cell entries are predicted values from Model 2 of Appendix Table A1. Standard errors in parentheses. Two-sided p-values for post-estimation difference-of-proportions test in brackets. Supply Treatment indicates whether or not the precinct received one of the two Supply conditions (Supply or Supply+Demand). Demand Treatment indicates whether or not the precinct received one of the two Demand conditions (Demand or Supply+Demand).

Another way to examine the results is to compare each treatment condition to the control. This approach may yield additional insights because the Supply+Demand condition is unique in

---

[15] The interaction term in the model in Table A1 also shows that the effect of each main treatment is not contingent on the presence of the other treatment.

its request that the precinct chair take *two* actions and may bestow greater intensity or urgency from the state party about this issue such that the two treatments together exert an especially powerful effect. Thus, in subsequent analyses, we evaluate each of the three experimental conditions (Supply, Demand, and Supply+Demand) independently against the control.

Figure 2 gives a graphical representation of this comparison for the raw experimental results without any controls or other statistical adjustments. Error bars in the figure represent the 95 percent confidence interval and spikes represent the 83 percent confidence intervals. Showing the 83 percent confidence intervals allows viewers to easily compare across conditions and determine whether the differences between the point estimates are statistically significant at roughly the 95 percent confidence level.[16] When compared to the control, the Supply and Demand conditions on their own each fall short of standard levels of statistical significance. However, the Supply+Demand condition statistically significantly reduced the proportion of precincts that did not elect any women. In the control condition, only 37.5 percent of precincts elected at least one woman; this increases to 45.4 percent in the Supply+Demand condition.

---

[16] Visual examination of whether 95 percent confidence intervals overlap will result in Type II errors and is not a reliable way of testing differences in point estimates (see Bolsen and Thornton 2014 for an explanation and examples). Instead, 83 percent confidence intervals offer a rough visual approximation of a formal hypothesis test of a two-tailed difference of means test at the 95 percent confidence level (see Bolsen and Thornton 2014; Goldstein and Healy 1995; Payton, Greenstone, and Schenker 2003).
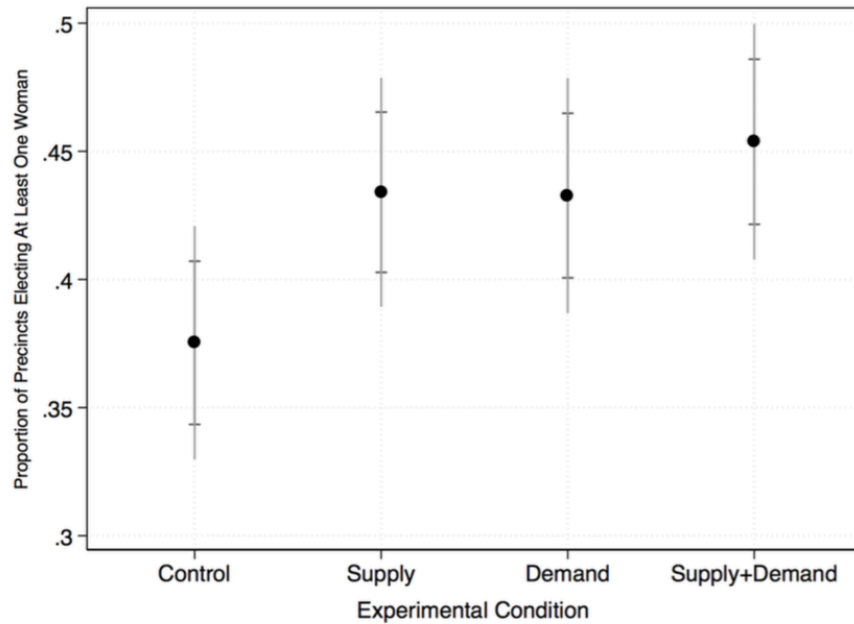
Note: Raw experimental results with no controls or adjustments for county.
Lines represent 95% confidence intervals. Spikes represent 83% confidence intervals.

Table 5 offers a formal test of the intent-to-treat effects of the treatments. Because the excluded category in the regression models is the control condition, the coefficients for the other treatments show the difference from the control. The state party relies heavily on county-level party officials to organize and run the neighborhood caucus meetings, so standard errors are clustered by county. All treatments have a positive coefficient, but only the Supply+Demand precincts saw a statistically robust change.

These results are sharpened by the inclusion of control variables, such as the number of delegate positions available (which varied with the size of the precinct and its tendency to vote Republican in previous elections), the average age of caucus participants who attended the meeting, the size of the meeting, and the distance between the precinct and the convention site (which was held in a suburb of the state's largest city). In addition, including the control variables reveals that a woman is more likely to be elected when a greater proportion of those at

the meeting were women, but the experimental treatments exerted a positive effect on women's representation independent of that relationship.[17] Across all conditions, attendance at meetings tends to come close to gender balance: about 45 percent of participants at the neighborhood gatherings were women, and the proportion of women among participants is roughly normally distributed (see Figure 1 and Supporting Information Figure A1). These patterns also reassure us that the failure to elect women is not merely the result of women being under-represented at the meetings. Further, the effects of the experimental conditions – especially the Supply+Demand treatment – persist even when we add an additional control for the precinct's 2012 proportion of female delegates (see Supporting Information Table A2).[18] In other words, even when we control for women's electoral success in the precinct at the previous caucus meeting, the Demand treatment and Supply+Demand treatment had a substantial effect on whether or not the precinct elected at least one woman in 2014.[19]

---

[17] Importantly, the experimental conditions had no effect on the proportion of caucus participants who were women. Whatever other effects they might have had, our interventions did not substantially alter the mix of men and women who attended the meeting. We can thus include that variable in the regression model without worrying about correlation between it and the experimental conditions.

[18] Here and in subsequent analyses, the overall number of observations is lower when we control for the 2012 results because some new precincts were created between 2012 and 2014, and we do not have a 2012 measure of women's electoral success in these new precincts.

[19] When we restrict the analysis to those precincts for which we had a working mailing address, the results are further strengthened (see Supporting Information Table A3), with the Supply+Demand condition increasing the proportion of precincts with at least one woman in the delegation by nearly 10 percentage points in models with controls.

Table 5: Intent-to-Treat Effects – Precincts Electing At Least One Woman

	OLS		Probit	
	(1)	(2)	(3)	(4)
Supply Condition	0.059	0.051	0.152	0.151
	(0.033)	(0.033)	(0.085)	(0.096)
	[0.085]	[0.127]	[0.075]	[0.117]
Demand Condition	0.057	0.053	0.148	0.147
	(0.036)	(0.030)	(0.092)	(0.089)
	[0.118]	[0.089]	[0.105]	[0.099]
Supply + Demand Condition	0.078	0.091	0.202	0.261
	(0.026)	(0.021)	(0.063)	(0.055)
	[0.005]	[0.000]	[0.001]	[0.000]
Number of Delegate Positions Available		0.174		0.533
		(0.026)		(0.075)
		[0.000]		[0.000]
Mean Age of Caucus Participants		0.001		0.003
		(0.001)		(0.005)
		[0.437]		[0.489]
Number of Caucus Participants		-0.001		-0.004
		(0.002)		(0.005)
		[0.421]		[0.389]
Proportion of Caucus Participants Who Are Women		1.038		3.520
		(0.088)		(0.324)
		[0.000]		[0.000]
Distance between Precinct and Convention Site		-0.007		-0.022
		(0.016)		(0.049)
		[0.669]		[0.650]
Constant	0.375	0.454	-0.318	-3.056
	(0.039)	(0.092)	(0.103)	(0.339)
	[0.000]	[0.000]	[0.002]	[0.000]
Observations	1,812	1,730	1,812	1,730
R-squared/Pseudo R-squared	0.004	0.178	0.003	0.149

Note: Dependent variable is a dummy variable indicating whether or not the precinct elected 1 or more women. Cluster robust standard errors in parentheses; cluster by county. Two-sided p-values in brackets. Excluded category is the control condition.

Proportion of State Delegates Who Are Women

Table 6 shows the intent-to-treat results for the overall proportion of the precinct's state

delegates who are women, comparing each of the treatments to the control.[20] Because the

---

[20] The analysis of the main treatment arms for this dependent variable can be seen in Supporting Information Table A4.

dependent variable is highly right-skewed (most precincts elected no women at all) and because

proportions are constrained by definition between 0 and 1, we model the relationships with both

OLS regression and fractional logit, which is a generalized linear model with a binomial family

and logit link. The key substantive results are nearly identical no matter which modeling

strategy we employ, whether OLS, fractional logit, zero one inflated beta, or tobit. In the results

that follow, whenever we compute predicted values, we use fractional logit models to avoid the

problem of predicted values falling outside the range of the dependent variable. The results in

Table 6 show that while both the Supply and the Demand condition have positive coefficients,

neither achieves statistical significance. However, the Supply+Demand condition does. Model 1

indicates that it increased the proportion of women elected about four percentage points; Model 2

includes controls for precinct characteristics and increases that effect size to 5.7 percentage

points.

Table 6: Intent-to-Treat Effects – Proportion of State Delegates Who Are Women

	OLS		Fractional Logit	
	(1)	(2)	(3)	(4)
Supply Condition	0.015	0.016	0.079	0.098
	(0.015)	(0.017)	(0.083)	(0.099)
	[0.343]	[0.363]	[0.340]	[0.322]
Demand Condition	0.015	0.023	0.079	0.137
	(0.021)	(0.021)	(0.113)	(0.125)
	[0.489]	[0.292]	[0.484]	[0.272]
Supply + Demand Condition	0.040	0.057	0.203	0.317
	(0.016)	(0.013)	(0.081)	(0.072)
	[0.019]	[0.000]	[0.012]	[0.000]
Number of Delegate Positions Available		0.031		0.179
		(0.011)		(0.066)
		[0.011]		[0.006]
Mean Age of Caucus Participants		0.001		0.008
		(0.001)		(0.005)
		[0.157]		[0.153]
Number of Caucus Participants		-0.002		-0.008
		(0.001)		(0.006)
		[0.108]		[0.174]
Proportion of Caucus Participants Who Are Women		0.953		5.700
		(0.066)		(0.442)
		[0.000]		[0.000]
Distance between Precinct and Convention Site		-0.015		-0.091
		(0.011)		(0.057)
		[0.157]		[0.106]
Constant	0.246	-0.267	-1.122	-4.330
	(0.015)	(0.065)	(0.079)	(0.405)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1,812	1,730	1,812	1,730
R-squared	0.002	0.112		
AIC			1.025	0.950
BIC			-12171.69	-11647.76
Log-pseudolikelihood			-924.23	-813.05

Note: Dependent variable is proportion of elected state delegates who are women. Cluster robust standard errors in parentheses; cluster by county. Two-sided p-values in brackets. Excluded category is the control condition.

Because fractional logit coeffcients are difficult to interpret, Figure 3 shows the predicted values for the proportion of state delegates who are women based on Model 4.[21] In the Control

---

[21] Estimates include Bonferroni corrections for multiple comparisons.

condition, women comprise less than 24 percent of state delegates, but that number rises to close

to 30 percent in the Supply+Demand condition, a difference that is significant at p<0.01.[22] In

Model 4, the estimated proportion of women in the Supply+Demand condition is also larger than

the proportion in the Supply alone condition (p=0.02). Finally, the estimated value of the

Supply+Demand condition is greater than that of the Demand condition alone, but the difference

in the estimated values falls short significance.[23] As with the previous dependent variable, these

results are robust to the inclusion of controls for women's electoral success in the precinct in

2012 (see Supporting Information Table A5) and to conditioning on those precincts for which a

letter was actually sent (Supporting Information Table A6).

The size of the treatment effects, relative to the control, is quite large – in field

experiments on voter turnout, for example, an intent-to-treat effect of 5 to 10 percentage points is

considered very large. In this case, the downstream mass effects of encouraging party elites to

take actions that favor women result in a 5 to 10 percentage point change in the gender balance

of the state's Republican convention. The substantive significance of this change may seem

modest at first glance but is more powerfully understood in raw numbers. If all precincts had

been assigned to the control condition, our analysis predicts that about 920 of the 3839 delegates

would have been women (a proportion also consistent with recent state conventions). In

contrast, if all precincts had been assigned to the Supply+Demand condition, over 1150 of the

delegates would have been women—an increase of 230 additional women attending the

convention or 25 percent more than the 920 that would have attended without any intervention.

---

[22] P-values are computed using a Wald test on predicted values from Model 4 of Table 6.

[23] All of these comparisons are estimated from Model 4 of Table 6. When we predict values Models 1-3, only the difference between the control and Supply+Demand condition is statistically significant at conventional levels.

Our experiment suggests that a simple intervention in a single state can bring hundreds of women into the formal political party structure and its associated informal networks.

Figure 3: Estimated Proportion of State Delegates Who Are Women, by Condition
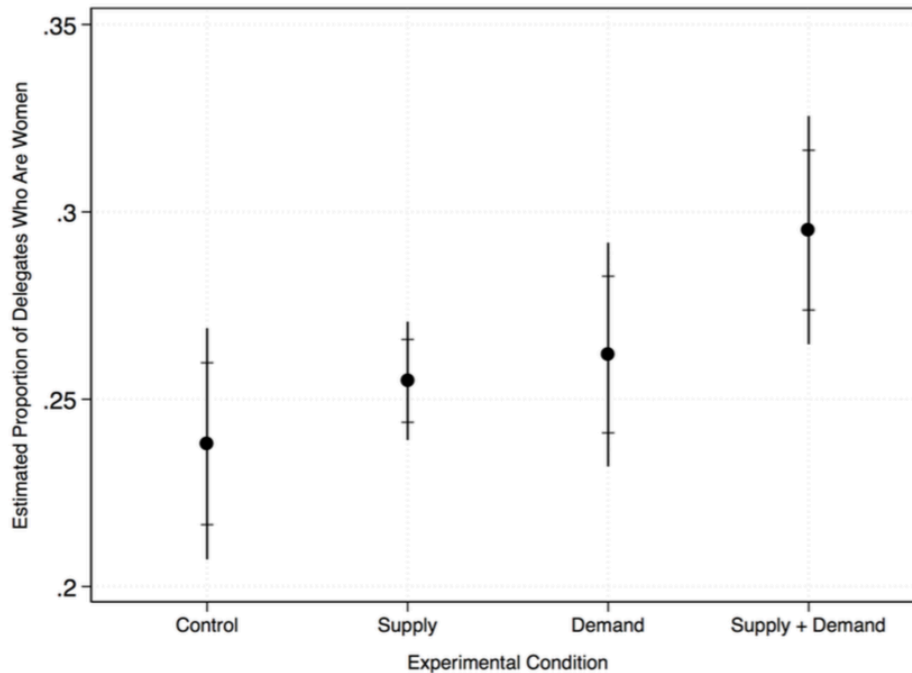
Predicted values from Model 4 in Table 6, with Bonferroni correction for multiple comparisons.
Lines represent 95% confidence intervals. Spikes represent 83% confidence intervals.

Finally, we also explored the effects on the overall proportion of women elected when dividing by one other pre-treatment variable: the gender of the precinct chair (elected two years prior) to whom the letter was sent. Table 7 shows that the effect of the Supply+Demand treatment was concentrated in precincts with a male precinct chair. This result is partly because the problem of women's under-representation is greatest in those precincts.[24] Precincts with a female chair were far less likely to elect an all-male slate of state delegates, and those precincts may have been distinct in other ways as well. Our experiment is not designed to isolate the cause

---

[24] See Supporting Information Figure A2 for a graphical presentation of the distribution of the proportion of women elected by the gender of the precinct chair to whom the letter was sent. The distributions across the two panels are very different; precincts with a male chair were far more likely to elect no women as state delegates.

of this particular difference. It might occur because the encouragement to vote for more women is especially efficacious when given by a male authority figure or because the problem of underrepresentation of women seems less acute when the encouragement is delivered by a female precinct chair, or some combination of these and other related factors.

Table 7: Intent-to-Treat Effects by Gender of Precinct Chair

| | Male Precinct Chair | | Female Precinct Chair | |
	(1)	(2)	(3)	(4)
Supply Condition	0.006	0.034	0.057	0.254
	(0.027)	(0.154)	(0.054)	(0.238)
	[0.825]	[0.823]	[0.297]	[0.284]
Demand Condition	0.013	0.074	0.048	0.215
	(0.024)	(0.136)	(0.087)	(0.385)
	[0.590]	[0.587]	[0.590]	[0.576]
Supply + Demand Condition	0.052	0.279	0.016	0.071
	(0.016)	(0.083)	(0.046)	(0.211)
	[0.004]	[0.001]	[0.739]	[0.735]
Constant	0.169	-1.266	0.313	-0.276
	(0.018)	(0.109)	(0.028)	(0.156)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1,324	1,324	410	410
R-squared	0.006		0.004	
AIC		0.972		1.198
BIC		-8534.956		-2708.51
Log-pseudolikelihood		-639.453		-300.29

Note: Dependent variable is proportion of elected state delegates who are women. Models 1 and 3 estimated with OLS; models 2 and 4 with fractional logit. Cluster robust standard errors in parentheses; cluster by county. Two-sided p-values in brackets. Excluded category is the control condition.

Mechanisms for Downstream Effects

Our focus so far has been the treatments' downstream effects on women's descriptive representation at the state party convention. This is ultimately the dependent variable that matters for party leaders' ability to remedy substantial gender imbalances. However, our theory held that these changes hinged on two specific mechanisms: an increase in the number of women running (supply) and a change in precinct-level norms about women's participation in the

nominating process (demand). If our theoretical approach is correct, we should also be able to see evidence of these mechanisms at work.

Evidence that the Demand Treatment Provoked a Change in Norms

Our evidence for the relationship between the experimental treatments and the precinct-level norms is drawn from two sources: 1) the 2014 Survey of Precinct Chairs and 2) a sub-sample of 145 precincts to which we sent trained observers.

In our survey, we asked the chairs about their attitudes toward women's participation in politics, including whether "more women are needed" as delegates to the state convention.[25] If our theory about norm-setting is correct, we should see different reported attitudes about the role of women among those who received the Demand message. Table 8 shows that precinct chairs in both the Demand and the Supply+Demand conditions express greater agreement with the idea that women should have increased presence at the convention, compared to the control. Precinct chairs in the Supply condition are no different from the Control.

As the table shows, the effect is concentrated among men: they are the ones who are moved by the party message, while women are not affected. A formal test comparing the effect of the Demand treatment in all the conditions that received it (Demand and Supply+Demand) to the conditions that did not include the Demand message (Supply and Control) shows that the effect is modest but robust.[26] Male precinct chairs who were asked to read a paragraph encouraging caucus-goers to elect more women had, on average, attitudes that were about .22 points friendlier to women on the 5-point response scale (3.36, compared to 3.14) than precinct chairs who were not randomly assigned to a condition with the Demand treatment (t=2.75,

---

[25] The exact statement was "More women are needed as convention delegates." Responses were given on a 5-point Likert scale, with higher scores reflecting increased agreement with the statement.

[26] This test is a simple difference-of-means test and does not include controls for other demographic characteristics. When such controls are included, the effect holds -- if anything, it is slightly larger and more statistically robust.

p<0.01). This represents a difference of about 4.4 percent of the full span of the dependent variable.

Table 8: Precinct Chair Attitudes about Women by Experimental Condition

	All	All	Men Only	Women Only
"More women are needed as convention delegates." (1-5)				
Supply Condition	0.023	0.068	0.103	0.043
	(0.074)	(0.077)	(0.104)	(0.274)
	[0.761]	[0.384]	[0.334]	[0.877]
Demand Condition	0.172	0.206	0.322	-0.079
	(0.081)	(0.120)	(0.171)	(0.203)
	[0.043]	[0.098]	[0.072]	[0.703]
Supply + Demand Condition	0.095	0.149	0.280	-0.092
	(0.063)	(0.069)	(0.122)	(0.256)
	[0.141]	[0.000]	[0.032]	[0.723]
Constant	3.194	3.478	3.824	2.988
	(0.059)	(0.843)	(0.228)	(0.399)
	[0.000]	[0.000]	[0.000]	[0.000]
Demographic Controls	No	Yes	Yes	Yes
Observations	578	520	376	143
R-squared	0.006	0.022	0.057	0.015

Note: Table entries are OLS regression coefficients with cluster robust standard errors in parentheses; cluster by county. Two-sided p-values in brackets. Excluded category is the control condition. Demographic controls include income, age, religion, and race/ethnicity. Full models with controls available in Supporting Information Table A8.
Source: 2014 Survey of Precinct Chairs.

These effects are relatively small, and the data are limited because they are only self-reports from the subset of precinct chairs who responded to the survey, not measures of attitudes among caucus-attenders more generally. Nonetheless, the differences are consistent with the belief that persuasive messages from party elites can affect norms for women's role in politics. Even if the difference is entirely attributed to social desirability it would indicate some shift in norms. In addition, the change persisted for several weeks after the meeting until the survey of precinct chairs was fielded. That our one-time treatment had any discernible effect at all several weeks after the conclusion is the caucus meetings is worthy of note.

The precinct chair survey data are also supplemented by one additional piece of evidence. If norms about the importance of women's voices being heard are changed as a result of our experimental treatment, we might also expect that the women who stand for election as delegates will be more assertive and talkative during the meeting (Karpowitz and Mendelberg 2014). One way of measuring women's greater assertiveness is to examine the proportion of female candidates who gave speeches urging others to vote for them. At the caucus meeting, all candidates for office are given an opportunity to make a short speech about why they should be elected to the state convention, though not all candidates choose to speak up. We sent trained observers who were blind to the purposes of the study to a small convenience sample (n=145) of the precinct meetings.[27] We asked the observers to keep systematic notes about the proceedings and the candidates who ran for office, including which candidates gave speeches and which did not. This sample is not fully representative of the state; instead, it includes precincts within a reasonable driving distance of the students' campuses.[28] Nonetheless, we randomly sampled accessible precincts from each experimental condition, and the data from the student observers is the best precinct-level record we have of the speeches and the gender of candidates.

In precincts assigned to the read the Demand paragraph (the Demand and Supply+Demand conditions), women candidates spoke more often. Not only were a greater proportion of the precinct's speeches given by women (31percent, as compared to 23 percent, t=1.73, p=0.09, two-tailed precinct-level difference-of-means test), but a greater proportion of female candidates spoke up. There is no comparable effect for the Supply treatment. Overall,

---

[27] The experimental conditions were evenly distributed across the set of precinct meetings to which we sent observers.

[28] Students at three different university campuses in three different counties in the state participated in the project. They attended the neighborhood meetings in teams of two, and levels of agreement between the coders were very high. Coders agreed on the gender of the candidates in 99 percent of cases and they agreed on whether or not the candidate made a speech 95.5 percent of the time. For purposes of this analysis, the small number of cases where coders disagreed about either the gender of the candidate or the presence of a candidate speech were dropped.

nearly 90 percent of female candidates gave speeches in precincts where the chair was asked to read the Demand paragraph, compared to 78 percent of candidates in precincts that were not assigned to the Demand appeal (t=1.68, p=0.09). These results are only an indirect measure from a limited sample of precincts, but they are consistent with our theory that elite messages can change expectations about women's empowerment and active participation.

Evidence that the Supply Treatment Affected the Number of Women Running for Office

To support the other side of our theory, we need evidence that the Supply treatment had an effect on the number of women who stood for election. Based on the data in our precinct chair survey, about 57 percent of precinct chairs in the conditions that included the Supply treatment told us that they recruited at least one woman to run, as opposed to only 29 percent in the remaining conditions (z=5.56, p<0.01, two-tailed difference-of-proportions test). If those results are more than mere reporting bias, then we should see evidence in the actual number of women running. Unfortunately, the state party does not collect any information on the candidates who stand for election in each precinct; their formal reporting process includes only the identity of the electoral winners. For that reason, we turn again to our student observer data.

Models 1 and 2 of Table 9 show that even in our limited sample, the experimental conditions in which chairs were asked to recruit women (Supply and Supply + Demand) also had slightly higher numbers of female candidates, though that relationship is only statistically different from the control condition in the Supply + Demand condition.[29] Models 3 and 4 compare the two conditions with the recruitment message (Supply and Supply + Demand) with the two conditions that did not include such encouragement. Though many other factors beyond

---

[29] Because the dependent variable is a count that is skewed toward 0, we employ a negative binomial regression.

our experimental treatments are clearly at work, these models provide evidence that when the precinct chair was encouraged to recruit women, more women ran.

Table 9: Number of Female Candidates by Experimental Condition

	(1)	(2)	(3)	(4)	(5)
Supply Condition	0.081	0.105	–	–	–
	(0.257)	(0.251)			
	[0.754]	[0.675]			
Demand Condition	0.001	-0.064	–	–	–
	(0.262)	(0.257)			
	[0.998]	[0.803]			
Supply + Demand Condition	0.403	0.403	–	–	–
	(0.238)	(0.232)			
	[0.090]	[0.082]			
Supply Treatment	–	–	0.263	0.307	0.881
			(0.176)	(0.172)	(0.36)
			[0.135]	[0.074]	[0.014]
Number of Caucus Participants	–	0.013	–	0.013	0.005
		(0.004)		(0.004)	(0.009)
		[0.001]		[0.000]	[0.569]
Constant	0.027	-0.397	0.028	-0.433	-0.447
	(0.184)	(0.221)	(0.132)	(0.190)	(0.458)
	[0.881]	[0.073]	[0.833]	[0.023]	[0.330]
Log Alpha	-1.423	-1.740	-1.350	-1.683	-1.303
	(0.570)	(0.695)	(0.542)	(0.669)	(0.877)
Observations	145	145	145	145	38
Pseudo R-squared	0.01	0.04	0.01	0.03	0.05
Log likelihood	-219.39	-204.52	-211.33	-205.40	-56.72

Note: Dependent variable is a count of the number of women who ran for state delegate. Cell entries are negative binomial regression coefficients. Standard errors in parentheses. Two-sided p-values in brackets. Excluded category in models 1 and 2 is the control condition. Supply Treatment is a dummy variable for Supply and Supply + Demand conditions together.

How big is the effect? In the conditions without recruitment appeals, on average approximately one woman ran for state delegate in each precinct. When precincts were randomly assigned to the recruitment condition, that number rises to 1.34 – an increase of about 30 percent. That increase is slightly larger (about 36 percent) when controls for the size of the caucus meeting are included in the analysis. If we limit the analysis to the small number of

precincts where we know from our survey that the chair received the letter prior to the caucus meeting (Model 5 of Table 9), the effect is much larger still – a predicted increase of about 141 percent, with the estimated number of women running changing from less than 1 to nearly 2 (0.77 to 1.86). These results should be interpreted with caution because we have neither a random sample of observed precincts, nor full compliance information for that sample. Nonetheless, all of these results converge on the same, basic conclusion: when the precinct chair was encouraged by party leaders to increase the supply of female candidates, more women did, in fact, run.

Replication Vignette Experiment

Our field experiment reveals a robust relationship between political party intervention and increases in women's representation. This is the first study to show that an intervention from party leaders can increase women's representation, but like any research design, it also has limitations. First, it takes place in one state and focuses on internal political party offices, which could potentially limit generalizability. Second, it is difficult to untangle whether the Supply+Demand treatment was especially effective because the two mechanisms interact in important ways or because asking party chairs to do two tasks signaled extra seriousness from the party. And finally, spillovers and two-sided non-compliance potentially pose problems for causal inference.

We conducted a replication study to address these limitations and test the robustness of our field experimental findings. We designed a vignette experiment fielded by YouGov with a national sample of Republican primary election voters (N=2,897) in the 30 states for which

validated primary vote data *and* party registration data were available.[30] Because the vignette is about gender in a state legislative election, we block randomized by respondent gender and by whether the respondent was from a state above or below average in female representation among Republican state legislators.[31]

As in the field experiment, respondents were randomly assigned to one of four conditions: Control, Supply, Demand, and Supply+Demand. In the Control condition, respondents were asked to choose from four hypothetical candidates for a primary election for a state legislative seat in their state—three men and one woman. Figure 4 shows a screenshot of the control condition vignette. In addition to candidate names, which indicate the sex of the candidate, we also included a number of other traits to add a sense of realism to the vignette and mimic the kinds of information that a voter might find in a voter guide. These traits included age, occupation, political experience, and a short personal statement that reflected the candidate's ideology. The information in the candidate profiles was independently randomized on the following dimensions: first name, last name+age+occupation,[32] political experience, personal statement and order of profile appearance. Independently randomizing the characteristics allows us to isolate the causal effects of our experimental treatments and reassures us that those effects are not the result of any particular other characteristic or the order in which the profiles appeared.

---

[30] YouGov matched panelists against the state voter lists by name and address and selected respondents who were identified as voters registered with the Republican Party. Where available, YouGov selected respondents who had voted in Republican Primaries in the years 2012 or 2014 in states that reported this information, and registered Republicans who were verified to have voted in a primary otherwise.

[31] States above average in their proportion of female Republican legislators include: Alaska, Arizona, Arkansas, California, Colorado, Florida, Idaho, Illinois, Maine, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New Mexico, Ohio, and Pennsylvania. Those below average include: Alabama, Delaware, Georgia, Indiana, Iowa, Michigan, Mississippi, New York, North Carolina, South Carolina, Tennessee, Texas, Virginia, and Wyoming.

[32] Sample size constraints kept us from using a completely full factorial design, so we linked together last name, age, and occupation.

Figure 4: Screenshot of Control Condition Vignette

YouGov

Imagine the following slate of candidates is running in a **Republican primary election for your state legislature**. Please choose the candidate you would vote for in this situation:

Name	Patrick Sanders	Daniel Powell	David Jenkins	Christine Anderson
Age	50	47	52	39
Occupation	Physician	Small Business Owner	High School Teacher	Attorney
Political Experience	City councilor	Volunteer on three previous campaigns	School board member	Republican Party precinct chair
What do we need to do to improve the country?	Liberal big government is bankrupting the country. We need to end entitlements and return to the principles of constitutional government.	The government needs to get out of our lives and respect individual rights. Leave people alone and let them make their own decisions.	We need to come up with practical solutions to balance the budget and solve other problems. Partisan bickering is a distraction from that.	Businesses create jobs and drive the economy. Current economic policy is bad for business, and we need to reduce regulation and red tape to stimulate growth.

Which candidate would you vote for in a Republican primary election for your state legislature?

◯ Patrick Sanders
◯ Daniel Powell
◯ David Jenkins
◯ Christine Anderson

< >

In the Demand condition, respondents were presented text from the Republican National Committee's Growth and Opportunities Project report that encouraged Republican voters to elect more women, as well as excerpts from the letter that was sent with the field experiment (see Supporting Information XX for the full text). In the Supply condition, respondents chose between five (instead of four) hypothetical candidates—three men and two women. This additional woman is to simulate the result of active party recruitment of female candidates. In the Supply+Demand condition, respondents saw both the text from Republican leaders encouraging voters to elect more women and saw two (rather than one) female candidates on the ballot.

Our outcome of interest is whether the respondent chose a female candidate. In Figure 5, we present the percentage of respondents who chose a female candidate as increase (or decrease) over the percentage of respondents who would have chosen a female candidate if candidates were chosen at random. This approach allows us to directly compare the size of the effect across treatments. In the Control and Demand conditions, the baseline of random choice is 25 percent because there was one woman out of four candidates; in the Supply and Supply+Demand conditions, the baseline is 40 percent because there were two women out of five candidates.

Figure 5: Effects of Experimental Treatments in Sample of Republican Primary Voters
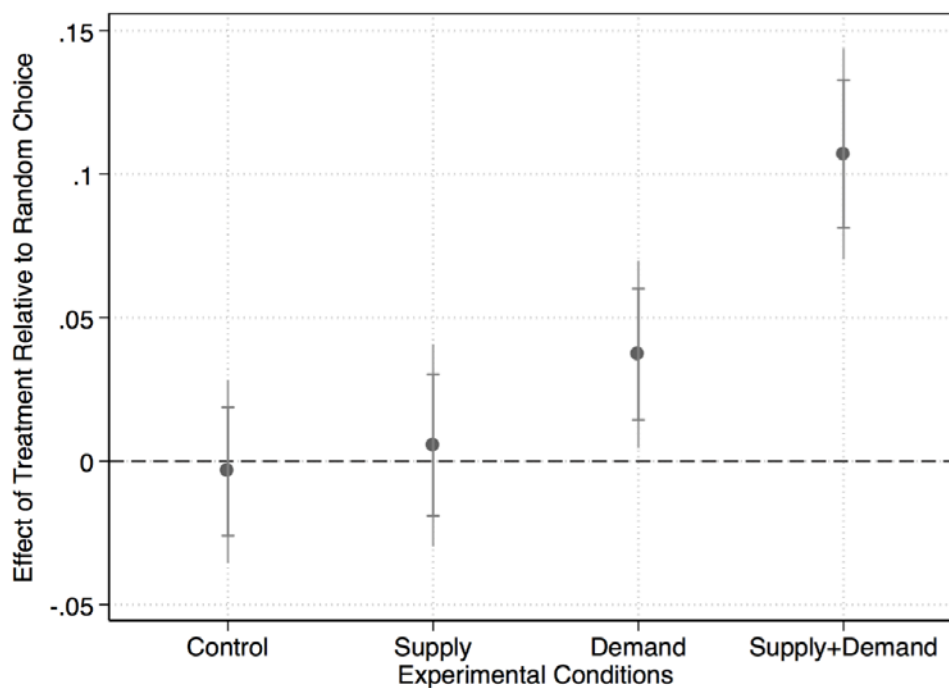
Predicted values from Model 2 in Table A9, with Bonferroni correction for multiple comparisons.
Lines represent 95% confidence intervals. Spikes represent 83% confidence intervals.

Figure 5 indicates that in the Control condition, respondents chose the female candidate at almost exactly the rate expected if the candidate's gender was irrelevant to decision-making. This finding lends support to previous research suggesting that, by and large, voters do not

discriminate against women on the ballot. Figure 5 also indicates that having additional women

on the ballot (Supply) may increase the chance that a woman is elected, but only because there

are more women on the ballot; voters continue to select candidates as if at random. However,

when Republican voters are urged to support female candidates by credible party sources

(Demand), female candidates do nearly 4 percentage points better than the baseline comparison

point. This is a significant finding—simply by emphasizing the importance of women's voices

in the party, Republican leaders are able to shape Republican voters' preferences for state

legislative candidates. More importantly, the Supply+Demand condition has an enormous

influence on respondents' vote choices. In this condition, voters were about 11 percentage points

more likely to choose a female candidate. Thus, party leaders' encouragements to vote for

women affect vote choice, but leaders' interventions paired with more female candidates on the

ballot is especially effective.

This experiment clearly replicates our findings from the field and provides reassurance

that those results are likely genuine for at least three reasons. First, while the field experiment

suffers from some generalizability limitations, the vignette experiment takes advantage of a large

national sample and studies a different political office. Second, in the field experiment, we were

unable to differentiate why Supply+Demand was so effective, but because respondents in the

vignette experiment were not aware that they experienced two interventions, there is, in fact,

something jointly important about the two interventions beyond increased intensity. Finally, the

field experiment may have had two-sided noncompliance posing problems for causal inference,

but the controlled nature of the vignette experiment prevented spillover between conditions. In

short, we are now more confident that, on their own, increasing the supply of female candidates

or encouraging voters to elect women may have a marginally positive effect on women's

representation. But, together, these kinds of interventions have a meaningful effect on the number of women who get elected.

Discussion and Conclusions

In summary, our experimental interventions targeted two potential reasons for the underrepresentation of women in politics: the supply of female candidates and voter demand for female representatives. These two mechanisms are rarely linked together in the literature. Our main theoretical contribution is that they should be because our results suggest that the combination of both is especially powerful. More substantively, while there have been a handful field experiments on political party recruitment and a number of survey experiments on voter stereotypes, our field experiment is the first to actually increase the number of women elected to meaningful political offices.

Our research design also provides a model for how to ethically study elections with field experiments. Humphreys (2011, 2014) argues that field experimentation raises concerns when researchers are "taking actions that may have major, direct, and possibly adverse effects on the lives of others" (Humphreys 2011, 23). Obtaining prior consent of precinct chairs would have compromised the internal validity of the experiment, but full partnership and approval of the state party officials along with the fact that the actions were only requested, not required, of anyone alleviates these concerns.[33] Using an encouragement design probably weakened the effects, but beyond preserving the autonomy of precinct chairs, the field experiment's treatments

---

[33] In open-ended responses to our survey, some precinct chairs said they did not recruit candidates and/or read the letter encouraging voters to elect women because they felt such actions would impinge on voters' preferences at the meeting.

reflect realism about what party officials who remain resistant to quotas are actually prepared to do.

Conducting ethical field experiments also requires researchers to weigh potential harm against net benefits. The treatment objective to elect more women did not promote any overt ideological objective, thus minimizing harm. Any harm to individual male candidates should be seen in light of larger benefits. The goal of electing more women as convention delegates was seen by party officials as extremely desirable and beneficial for the state party and a more descriptively representative outcome is also viewed favorably by society in general. Finally, it is also important to factor in the harm that results from women's stark under-representation in the absence of any intervention. In the end, the confluence of the party's own desire to increase women's representation and its willingness to cooperate resulted in research that served party needs, minimized harm, maximized benefit, and thus provided an opportunity for ethical field experimentation.[34]

Ultimately, the field and survey experiments together are most important for its practical implications: even political parties that are ideologically opposed to quotas or other stronger interventions have the ability to increase women's representation in meaningful ways. Although our interventions were mild—a single letter sent to precinct chairs with a request from the state party chair or exposing voters to statements from party leaders—we saw significant increases in women's representation. To this point, the literature has focused on barriers to women's representation, but these experiments are the first to examine interventions that lead to greater representation for women.

---

[34] We are not asserting that field experimentation should stop whenever shared interests with a partnering organization are absent, but rather that researchers should seize such moments of opportunity when they arise.

Our experiments have also focused on the context in which gender disparities are the most pronounced. The Democratic Party has seen steady increases in the percentage of women elected—women now make up about a third of Democratic state legislators. However, movement in representation by women among Republican legislators has stagnated over the past 30 years (Carroll and Sanbonmatsu 2013, 68). This pattern holds among Members of Congress (Thomsen 2015). Republicans thus present the biggest challenge to increasing women's representation. By taking simple steps to encourage local recruitment efforts and emphasize that women's representation is a priority, Republican Party leaders can produce change.

References

Bolsen, Toby, and Judd R. Thornton. 2014. "Overlapping Confidence Intervals and Null Hypothesis Testing." *The Experimental Political Scientist* 4(1): 12–16.

Broockman, David E. 2014. "Mobilizing Candidates: Political Actors Strategically Shape the Candidate Pool with Personal Appeals." *Journal of Experimental Political Science* 1(2): 104–19.

Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes Do Not Harm Women Candidates*. Princeton University Press. http://books.google.com/books?id=WRxvqCAX1rkC.

Brown, Clyde, Neil R. Heighberger, and Peter A. Shocket. 1993. "Gender-Based Differences in Perceptions of Male and Female City Council Candidates." *Women & Politics* 13(1): 1–17.

Burns, Nancy, Kay Lehman Schlozman, and Sidney Verba. 2001. *The Private Roots of Public Action: Gender, Equality, and Political Participation*. Harvard University Press.

Carroll, Susan J, and Kira Sanbonmatsu. 2013. *More Women Can Run: Gender and Pathways to the State Legislatures*. New York: Oxford University Press.

Clark, Janet, Charles D. Hadley, and R. Darcy. 1989. "Political Ambition Among Men and Women State Party Leaders: Testing the Countersocialization Perspective." *American Politics Research* 17(2): 194–207.

Crowder-Meyer, Melody. 2013. "Gendered Recruitment without Trying: How Local Party Recruiters Affect Women's Representation." *Politics & Gender* 9(04): 390–413.

Crowder-Meyer, Melody, and Benjamin E. Lauderdale. 2014. "A Partisan Gap in the Supply of Female Potential Candidates in the United States." *Research & Politics* 1(1). http://rap.sagepub.com/content/1/1/2053168014537230.abstract.

Dolan, Kathleen. 2010. "The Impact of Gender Stereotyped Evaluations on Support for Women Candidates." *Political Behavior* 32(1): 69–88.

Dolan, Kathleen, and Timothy Lynch. 2013. "It Takes a Survey: Understanding Gender Stereotypes, Abstract Attitudes, and Voting for Women Candidates." *American Politics Research*: 1532673X13503034.

Elder, Laurel. 2012. "The Partisan Gap Among Women State Legislators." *Journal of Women, Politics & Policy* 33(1): 65–85.

———. 2014. "Contrasting Party Dynamics: A Three Decade Analysis of the Representation of Democratic versus Republican Women State Legislators." *The Social Science Journal* 51(3): 377–85.

Fox, Richard L., and Jennifer L. Lawless. 2004. "Entering the Arena? Gender and the Decision to Run for Office." *American Journal of Political Science* 48(2): 264–80.

———. 2005. "To Run or Not to Run for Office: Explaining Nascent Political Ambition." *American Journal of Political Science* 49(3): 642–59.

———. 2010. "If Only They'd Ask: Gender, Recruitment, and Political Ambition." *The Journal of Politics* 72(2): 310–26.

———. 2011. "Gendered Perceptions and Political Candidacies: A Central Barrier to Women's Equality in Electoral Politics." *American Journal of Political Science* 55(1): 59–73.

Fox, Richard L., and Eric R. A. N. Smith. 1998. "The Role of Candidate Sex in Voter Decision-Making." *Political Psychology* 19(2): 405–19.

Fridkin, Kim L., Patrick J. Kenney, and Gina Serignese Woodall. 2009. "Bad for Men, Better for Women: The Impact of Stereotypes During Negative Campaigns." *Political Behavior* 31(1): 53–77.

Fulton, Sarah A., Cherie D. Maestas, L. Sandy Maisel, and Walter J. Stone. 2006. "The Sense of a Woman: Gender, Ambition, and the Decision to Run for Congress." *Political Research Quarterly* 59(2): 235–48.

Gerber, Alan S, and Donald P Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94(3): 653–63.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Goldstein, Harvey, and Michael J. R. Healy. 1995. "The Graphical Presentation of a Collection of Means." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158(1): 175–77.

Grose, Christian R. 2014. "Field Experimental Work on Political Institutions." *Annual Review of Political Science* 17(1): 355–70.

Huddy, Leonie, and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37(1): 119.

Humphreys, Macartan. 2011. "Ethical Challenges of Embedded Experimentation." *Compartative Democratization (APSA Newsletter)* 9(3): 10–29.

———. 2014. "How to Make Field Experiments More Ethical." *Monkey Cage*. http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/11/02/how-to-make-field-experiments-more-ethical/ (January 3, 2015).

Kanthak, Kristin, and Jonathan Woon. 2013. "Women Don't Run? Election Aversion and Candidate Entry." *University of Pittsburgh*.

Karpowitz, Christopher F., and Tali Mendelberg. 2014. *The Silent Sex: Gender, Deliberation, and Institutions*.

King, David C., and Richard E. Matland. 2003. "Sex and the Grand Old Party An Experimental Investigation of the Effect of Candidate Sex on Support for a Republican Candidate." *American Politics Research* 31(6): 595–612.

Koch, Jeffrey W. 2002. "Gender Stereotypes and Citizens' Impressions of House Candidates' Ideological Orientations." *American Journal of Political Science* 46(2): 453.

Krupnikov, Yanna, and NicholeM. Bauer. 2014. "The Relationship Between Campaign Negativity, Gender and Campaign Context." *Political Behavior* 36(1): 167–88.

Lawless, Jennifer L. 2004. "Women, War, and Winning Elections: Gender Stereotyping in the Post-September 11th Era." *Political Research Quarterly* 57(3): 479–90.

Lawless, Jennifer L., and Richard L. Fox. 2010. *It Still Takes a Candidate: Why Women Don't Run for Office*. Cambridge University Press.

Loewen, Peter John, Daniel Rubenson, and Leonard Wantchekon. 2010. "Help Me Help You: Conducting Field Experiments with Political Elites." *The ANNALS of the American Academy of Political and Social Science* 628(1): 165–75.

Lupia, Aurthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

Maestas, Cherie D, Sarah Fulton, L Sandy Maisel, and Walter J Stone. 2006. "When To Risk It? Institutions, Ambitions, and the Decision to Run for the US House." *American Political Science Review* 100(02): 195–208.

Mo, CeciliaHyunjung. 2014. "The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters." *Political Behavior*: 1–39.

Niven, David. 1998. *The Missing Majority: The Recruitment of Women as State Legislative Candidates*. Greenwood Publishing Group.

Noel, Hans. 2012. "The Coalition Merchants: The Ideological Roots of the Civil Rights Realignment." *The Journal of Politics* 74(01): 156–73.

Payton, Mark E., Matthew H. Greenstone, and Nathaniel Schenker. 2003. "Overlapping Confidence Intervals or Standard Error Intervals: What Do They Mean in Terms of Statistical Significance?" *Journal of Insect Science* 3. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC524673/ (February 21, 2015).

Sanbonmatsu, Kira. 2002. "Gender Stereotypes and Vote Choice." *American Journal of Political Science* 46(1): 20–34.

———. 2006a. "Do Parties Know That 'Women Win'? Party Leader Beliefs about Women's Electoral Chances." *Politics & Gender* 2(04): 431–50.

———. 2006b. *Where Women Run: Gender and Party in the American States*. University of Michigan Press.

Sanbonmatsu, Kira, and Kathleen Dolan. 2009. "Do Gender Stereotypes Transcend Party?" *Political Research Quarterly* 62(3): 485–94.

Sapiro, Virginia. 1982. "Private Costs of Public Commitments or Public Costs of Private Commitments? Family Roles versus Political Ambition." *American Journal of Political Science* 26(2): 265–79.

Thomsen, Danielle. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." *Journal of Politics* 76(3): 786–97.

———. 2015. "Why So Few (Republican) Women? Explaining the Partisan Imbalance of Women in the U.S. Congress." *Legislative Studies Quarterly* 40(2): 295–323.

Wängnerud, Lena. 2009. "Women in Parliaments: Descriptive and Substantive Representation." *Annual Review of Political Science* 12: 51–69.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge University Press.