Women Don't Run? Election Aversion and Candidate Entry

Kristin Kanthak   University of Pittsburgh
Jonathan Woon   University of Pittsburgh

To study gender differences in candidate emergence, we conduct a laboratory experiment in which we control the incentives potential candidates face, manipulate features of the electoral environment, and measure beliefs and preferences. We find that men and women are equally likely to volunteer when the representative is chosen randomly, but that women are less likely to become candidates when the representative is chosen by an election. This difference does not arise from disparities in abilities, risk aversion, or beliefs, but rather from the specific competitive and strategic context of campaigns and elections. Thus, we find evidence that women are election averse, whereas men are not. Election aversion persists with variations in the electoral environment, disappearing only when campaigns are both costless and completely truthful.

L egislatures in the United States and around the world are becoming increasingly diverse yet remain unrepresentative. From 1945 to 2012, the number of minority members of Congress rose from 3 to 90, and the number of women rose from 11 to 102. Nevertheless, Congress remains 83% white and 83% male. In the states, 89% of legislators are white and 86% are male. Similarly, the number of women in legislatures around the globe has increased fourfold since 1945, but men still hold 88% of worldwide legislative seats.

The lack of diversity in contemporary legislatures is a significant problem not only from the perspective of descriptive representation and identity politics, but also for democratic decision making in general. Diverse groups can make better decisions than homogeneous ones because diverse perspectives offer the potential for creative decisions (Page 2007). For example, groups with more women tend to work together more effectively

(Woolley et al. 2010). In politics, legislatures with more women enjoy greater legitimacy for their decisions (Schwindt-Bayer and Mishler 2005), women tend to be more effective legislators (Anzia and Berry 2011; Volden, Wiseman, and Wittmer 2013), and their presence mitigates the deleterious effects of ideological preference divergence (Kanthak and Krause 2010, 2012). To effectively represent constituents' interests requires legislators to have a variety of strengths, cognitive styles, and interpersonal skills. The diversity of legislatures is therefore central to representation properly understood. But why aren't legislative bodies more diverse?

Political scientists have identified many reasons for the lack of diversity. Voters' stereotypes (Huddy and Terkildsen 1993), partisan politics (Wolbrecht 2000), or an otherwise skewed electoral process might be to blame (Sanbonmatsu 2006). But we also know that when women run for office, they win with at least as much frequency as

do men (Darcy et al. 1994). Recent scholarship has therefore focused instead on the question of candidate emergence (Lawless and Fox 2005). The decision of whether to run comes down to a variety of considerations, including incumbent characteristics (Stone and Maisel 2003), the potential candidate's current position (Johnson, Oppenheimer, and Selin 2012), outside recruitment (Fox and Lawless 2010), district magnitude (Matland and Brown 1992), party politics (Matland and Studlar 1996), or political efficacy and ambition (Campbell and Wolbrecht 2006; Fox and Lawless 2011; Maestas et al. 2006). Furthermore, although gender quotas have been brought to bear to address the issue in many countries worldwide (Krook 2009), their ability to challenge the majority status of men has been limited (Hughes 2011).

We propose an additional, distinctly behavioral explanation: Women may be more *election averse* than men. Even if potential candidates have the same qualifications, harbor the same ambitions, face the same incentives, and confront the same unbiased voters and electoral institutions—in short, encouter identical decision problems—the fact that representatives are chosen by electoral means is enough to dissuade women from putting themselves forward as candidates. To be clear, we claim neither that such a behavioral difference is the exclusive cause of underrepresentation nor that it is in any way innate or intrinsic.[1] Rather, this behavioral difference constitutes a distinct and powerful contributing factor that complements existing explanations.

Importantly, it is extremely difficult to use observational or survey data to disentangle election aversion from other inputs to the decision-making process, such as the desire to hold office, confidence, efficacy, recruitment, family obligations, or expectations of bias. But we can overcome these inferential challenges using experimental methods.[2] By controlling the incentives that potential candidates face, ensuring their identities remain anonymous, manipulating features of the electoral environment, and measuring crucial factors such as ability, risk preferences, and beliefs, our design rules out many

[1] See, for example, Gneezy, Leonard, and List (2009), whose findings suggest different behavior in matrilineal societies than in societies in which men have traditionally played the role of leader.

[2] Experimental elections are well-trodden ground in political science, having been called to task in, for example, studies of candidate positioning (McKelvey and Ordeshook 1985; Morton 1993), minority representation (Gerber, Morton, and Rietz 1998), pivotal voting (Battaglini, Morton, and Palfrey 2010; Duffy and Tavits 2008), accountability (Landa and Duell Forthcoming; Woon 2012a), and turnout (Feddersen, Gailmard, and Sandroni 2009; Grosser and Schram 2006; Levine and Palfrey 2007). The laboratory therefore provides an ideal and familiar setting in which to test for behavioral differences in decision making (Woon 2012b).

alternative explanations. The strength of our study is its high internal validity.

The results indeed point to gender differences in election aversion. Both men and women volunteer to be the representatives of their groups at equal rates, and they are equally responsive to task ability, provided that the selection of the representative does not involve an election. However, when selection does involve an election, women's willingness to represent decreases substantially. Furthermore, we show that the decline in candidate entry cannot be attributed to differences in ability, confidence about relative ability, or risk aversion. Instead, our findings indicate twin concerns: Campaigns are at once too costly and too noisy affairs. Women's entry into the candidate pool increases only if we simultaneously guarantee that campaigns are completely truthful and eliminate the private costs of running for office.

Experimental Design

Do men and women make different choices about becoming candidates when faced with the identical decision problem? If we could control for a variety of external and preference-based factors, such as differences in ambition, efficacy, political socialization, familial responsibilities, access to campaign fundraising, and political networks, would we still find behavioral evidence of gender-based election aversion? And if we establish that there are fundamental behavioral differences, do they stem from differences in underlying abilities, confidence in those abilities, and risk preferences, or do such differences arise instead from the very nature of electoral competition?

Several features of the design are critical to answering these questions. First, the experiment revolves around an objective problem-solving task. Performance on the task serves as an observable and objective measure of underlying ability, which we can think of as the laboratory analogue of policymaking skill—in other words, the potential quality of the representative. Such underlying abilities are impossible to observe using nonexperimental data. Second, the monetary payment scheme ensures that all members of a group have common incentives to select the highest-ability member as their representative. The alignment of individual and group interests eliminates competitive pressures and other-regarding tendencies from the decision problem. It also controls for heterogeneity in the value of holding office that is likely to exist between individuals and abstracts away from other sources of political conflict. Third, we compare the decision to enter the pool of potential representatives

FIGURE 1 Addition Task Screen

under alternative group selection mechanisms, varying whether the mechanism involves an election as well as varying the features of the campaign environment while holding the payoff structure fixed. By experimentally manipulating the selection mechanism, we carefully assess whether gender differences arise from differences in relative confidence in ability or depend on specific aspects of electoral competition. Fourth, we use a set of additional incentivized tasks to measure subjects' beliefs and risk preferences, which are potential sources of between-gender heterogeneity that would otherwise be unobservable. In all, our design takes advantage of many of the benefits of laboratory experimentation to strengthen the internal validity of our research findings.[3]

Addition Task

As the task for our experiment, we selected the Five-Minute Addition Task used by Niederle and Vesterlund (2007) to study preferences for competition. It involves computing the sum of five randomly selected two-digit numbers and doing as many of these sums as possible correctly within five minutes. For our purposes, the addition task has several desirable properties. It is specifically void of ideological or political content, which controls for subjects' knowledge and interest in particular political questions of the day. It is also a task for which there is heterogeneity between subjects and that previous research suggests is gender neutral. Figure 1 illustrates the

computer interface for the task, which we implemented in z-tree (Fischbacher 2007).[4]

Even though previous studies using the addition task demonstrate there are no gender differences in performance, the fact that the task involves doing math problems raises the possibility that stereotypes or stereotype threat—the concept that cuing on gender may decrease performance on the gendered task (Spencer, Steele, and Quinn 1999; Steele and Aronson 1995)—may play a role in the decisions of interest. That is, if negative stereotypes are deeply held, they might adversely influence women's abilities to perform the task well and therefore their decisions about representing their group. In our view, this is actually a desirable feature of our experimental design for the simple reason that politics, like math, is traditionally viewed as a task that belongs in the masculine domain (Conway, Steurnagel, and Ahern 1997) and therefore provides us with a harder test of election aversion than if we were to use a gender-neutral or feminine task. Nevertheless, we took precautions to guard against this (while also being careful not to specifically cue gender) by informing subjects that the "task has been chosen because there are no differences based on education level, socio-economic status, gender, or race in the ability of people to perform the task well." Indeed, Spencer, Steele, and Quinn (1999) demonstrate that simple cues such as this are effective at removing stereotype threat.[5]

Procedures

We conducted our experiments using the Pittsburgh Experimental Economics Laboratory at the University of Pittsburgh. A total of 350 subjects (173 men and 177 women) participated in the experiment. In each session of the experiment, we aimed to recruit 20 participants (equally divided between men and women), and each of these sessions lasted about an hour.[6] All interaction

---

[3]For the purposes of minimizing unobserved heterogeneity and enhancing internal validity, the use of a student sample is also an important feature of the experimental design: Undergraduates are at similar life stages, not yet having embarked on their careers or started their families, and their youth and education should also make them less susceptible to gender-based social constraints on running for office. A related concern, external validity, is "whether the causal relationship holds over variation in persons, settings, treatment variables, and measurement variables" (Shadish, Cook, and Campbell 2002, 38), but as many authors note, assessing external validity requires replication across different populations and contexts within a sustained program of research (Aronson and Carlsmith 1990; Druckman and Kam 2011; McDermott 2011; Morton and Williams 2010).

[4]After a subject enters a sum in the computer, the computer immediately presents the next series of random numbers and simultaneously provides feedback about whether the previous sum was correct, as well as a running tally of the number of correct sums. Subjects were not allowed to use calculators, but they could use scratch paper to complete the task.

[5]We address these issues in greater detail in the theoretical analysis. In related work, using an anagram task that is typically perceived as gender neutral, we find entry patterns similar to those reported here.

[6]See the supporting information for details about sample characteristics, randomization checks, and session information, as well as the full text of the instructions. We met our balanced gender recruitment target for 14 out of 18 sessions, and (as reported in the supporting information) our conclusions are robust to the exclusion of the unbalanced sessions from the analysis.

between subjects took place anonymously via a computerized interface, thus mitigating any potential effects of women's perceptions that voters may be biased against them. Subjects were assigned ID numbers and were identified in their interactions with other group members only by their ID number. Although subjects could observe the gender of the other participants in the session, we randomly assigned them to four groups of five members so that they would know neither which of the other participants were in their group nor their group's exact gender composition.

The procedures were divided into five parts, and at the beginning of each part we distributed and read the instructions aloud for that part of the experiment (so that anticipation of the later part of the experiment would not influence choices in the prior parts). At the end of the experimental session, subjects completed a questionnaire that included demographic questions, one of the five parts was randomly selected for payment (to guard against subjects using one part to hedge against decisions in other parts), and subjects were paid privately in cash. In addition to their earnings from one of the parts of the experiment, subjects received a $7 "show-up fee."

In Part 1, we introduced the addition task and incentivized subjects' performance using a simple *Piece Rate* compensation scheme. Each subject earns 75 cents for each correct sum if this part is selected for payment and receives feedback only about his or her own individual performance. No subject learns anything about the performance of the other members of the group. The purpose of Part 1 is for subjects to learn their ability in absolute, but not relative, terms.

In Part 2, we introduce *Group Representation*. Subjects first decide whether they are willing to be selected as the group representative, and then a representative is randomly selected from the set of willing members of each group.[7] We deem such willing members *volunteers* (although we do not use the term in our instructions so as not to induce or activate social desirability, norms, or other-regarding preferences). Subjects then repeat the task and, if Part 2 is selected for payment, are paid for their own performance plus the performance of the representative according to one of two incentive schemes. (The incentive scheme is the same for all subjects within a session and is thus a between-session manipulation.)

The two incentive schemes for Part 2 differ only in terms of whether they include an additional private cost

and benefit of volunteering for office. In the first variant (denoted *VNO*), subjects are paid 50 cents for each of the representative's correct answers plus 25 cents for each of their own correct answers.[8] Note that subjects maximize their payoffs (both individually and collectively) if the highest performer in the group is selected as the representative, and thus group members' preferences are aligned. As we will explain in the theoretical analysis, the decision to volunteer should primarily depend on subjects' *beliefs about relative ability*, and therefore Part 2 serves as our nonelectoral baseline for comparison.

In the second variant of the payment scheme for Part 2 (denoted *VCB*), we introduce additional costs and benefits. In addition to the 50 cents for each of the representative's correct answers and 25 cents for each of their own, volunteers earn an additional $2 bonus for being selected as the representative and pay a cost of $1 for volunteering (regardless of whether or not she was selected). These represent the private costs and benefits associated with participating in the selection process that are distinct from the common benefits associated with selecting a high-quality representative.

Part 3 is identical to Part 2 except that we replace the random selection mechanism with an *Election* for selecting the group representative. As in Part 2, subjects first choose whether they want to be considered to be the representative—that is, they choose whether or not to become a *candidate*. Candidates then engage in a brief "campaign" and every group member votes for one of the candidates; voting for oneself is permitted and no abstentions are allowed. The group's representative is then selected by plurality rule, with ties broken randomly. After the election, subjects repeat the addition task and are paid for performance on the task the same way as in Part 2 (depending on the session, either *VNO* or *VCB*). The incentives for selecting the highest performer are therefore the same as in Part 2, and the only difference is the selection mechanism.

If there are two or more candidates in their group, subjects engage in one of two possible kinds of campaigns. The type of campaign is a between-session manipulation and is therefore the same for all subjects within a session. In a *Chat* campaign, each candidate writes a brief text message (no more than 150 characters), and this message is the only information that other group members have when they vote. Candidates can write whatever they like, so the message is "cheap talk." Indeed, if subjects choose to do so, they can lie about their past performance on

---

[7]If no group member is willing, we randomly select one member from all members of the group.

[8]Although we provide feedback about subjects' own performance on the task in Part 2, we do not provide feedback about payoffs or the representative's performance until the end of the experiment so that subjects do not gain any information about relative abilities.

the task. In a *Truth* campaign, subjects do not actively send messages. Instead, every candidate's Part 1 score is revealed to the group, and this is the only information group members have when they vote. Thus, we compare a campaign environment that allows for the possibility of misinformation and strategic communication against one in which the only information voters have is truthful and payoff relevant.

Figure 2 clarifies the design and summarizes the three experimental manipulations: whether or not the selection mechanism involves an election, whether or not there are additional costs and benefits of participating in the selection process, and whether the election allows candidates to be free to campaign or if voters are automatically provided truthful information about candidate quality. The first manipulation provides for a within-subject comparison of volunteering in Part 2 with candidate entry in Part 3. The latter two manipulations imply a 2×2 factorial design that provides for between-subject comparisons. The resulting four "treatments" of the design are chat with costs and benefits (*CCB*), chat without costs and benefits (*CNO*), truth with costs and benefits (*TCB*), and truth without costs and benefits (*TNO*).

Part 4, which we called *Estimation*, involves an incentivized measurement task in which we elicited subjects' beliefs about other group members' performance and entry decisions. This task is important for assessing whether gender differences may be due to otherwise unobserved heterogeneity in relative confidence between men and women. We asked subjects to guess the Part 1 scores of the other four members of their group by rank order (highest, second highest, etc.) and rewarded them for their accuracy. In addition to asking about scores, we also asked subjects to guess what the other members' decisions were—to volunteer or to run—in Parts 2 and 3 and paid them for each decision guessed correctly. To guard against hedging, if Part 4 was selected for payment, we randomly selected a set of guesses corresponding to only one of the other members for payment.[9]

Our method of belief elicitation provides us with enough information to compute subjects' beliefs about the abilities of the pool of volunteers and about the pool of candidates by combining information about their guesses of Part 1 scores with their guesses about the Parts 2 and 3 decisions. Similarly, it is straightforward to use a subject's own task performance to determine how she thinks she measures up in comparison with the other subjects in her group; that is, we can use a subject's actual score and her elicited beliefs to assess relative confidence.

In Part 5, we measured risk preferences with an incentivized *Lottery Choice* task. The choice task we designed is similar to Holt and Laury (2002) except that we tailored the lotteries to correspond to potential payoffs from Part 2 of the experiment (random selection without costs and benefits). We presented subjects with a series of nine binary choices between a riskier lottery (corresponding to not volunteering) and a less risky lottery (corresponding to volunteering).[10] Because the latter gamble is less risky, subjects who exhibit greater risk aversion will require higher expected values of the risky option before switching; thus, the number of times a subject chooses the safer option is a measure of risk aversion.

Theoretical Analysis

To understand how our experimental design allows us to draw careful inferences about the various factors that might affect the decision to represent one's group, it is crucial that we analyze the incentives that group members face. By doing so, we can form clear expectations regarding the ways in which alternative assumptions would lead to observable differences in behavior so that we may be able to properly interpret our results. The starting point for the analysis is to consider the decision from the perspective of a risk-neutral, payoff-maximing individual. This provides us with a gender-neutral benchmark for behavior. We then consider how gender differences in abilities, beliefs about ability, beliefs about electoral competition, and risk preferences would lead to different patterns of behavior under alternative selection mechanisms.

Volunteers: Ability and Confidence

The volunteering decision provides us with a benchmark case in which all subjects share the common goal of selecting the best representative and in which election-specific factors are absent. Under the random selection mechanism without costs and benefits (*VNO*), an individual's decision comes down to a comparison of only two factors: her own *score* and her subjective *beliefs* about other volunteers' scores. If she is risk neutral, she will volunteer if she believes her score s_i to be above the average

[10]More specifically, to construct the lotteries, consider a situation in which a group member with a score of 10 believes that two other group members with scores of x and $x + 10$ have volunteered. Each binary choice corresponds to an integer value of x between 1 and 9. As x increases, the expected value of not volunteering increases. See the Measurement Appendix in the supporting information for further details.

FIGURE 2 Summary of Experimental Design

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Part 1: Piece Rate	Same in all treatments			
Part 2: Group Representation	Cost (VCB)		Vol. + Cost (VNO)	
Part 3: Election	Chat + Cost (CCB)	Truth + Cost (TCB)	Chat + No Cost (CNO)	Truth + No Cost (TNO)
Part 4: Estimation	Same in all treatments			
Part 5: Lottery Choice	Same in all treatments			
Number of Subjects	80	90	100	80
Men	40	43	50	40
Women	40	47	50	40

score of the other volunteers: $s_i \geq E[v_j]$, where $E[v_j]$ represents her belief about the average score of other volunteers $j \neq i$. The intuition is straightforward.[11] Given random selection, each volunteer is equally likely to be selected as the group's representative. If her score is above $E[v_j]$, then volunteering raises the expected score of the group's representative, but if her score is below $E[v_j]$, then volunteering lowers it.

This simple decision rule has direct implications for the conditions under which we would observe gender differences in candidate entry. If there are no differences in performance *and* if men and women have identical beliefs about the pool of volunteers, then we would not expect to observe differences in volunteering decisions. Conversely, we would expect to observe differences under either of two conditions. If beliefs are the same but there are *differences in performance* (i.e., men tend to do better on the task), then men would be more likely to volunteer than women, essentially corresponding to a scenario in which men are objectively "more qualified." Alternatively, if task abilities are the same but there are *differences in beliefs* (i.e., men believe others' scores are lower than women believe them to be), then men in this scenario would also be more likely to volunteer than women—a scenario in which men are "overconfident" and women "underconfident."

We chose the addition task in part because it has the potential to induce differences in beliefs. If the simple fact that the task involves math primes negative stereotypes that women hold about their math abilities, then they may systematically believe themselves to be "less qualified" even if they have the same ability. Our design specifically tests this possibility in several ways. First, if this stereotype explanation for gender differences is correct, then when we control for ability, we should see that women are less likely than men to volunteer *absent* electoral considerations—that is, in the *VNO* and *VCB* conditions. Second, we should observe that men and women give systematically different responses in our belief elicitation task, with women believing that the average score of other group members is higher than what men believe it to be. Third, if stereotypical beliefs account for women's candidate entry decisions, then we should not observe any change in the entry gap when we manipulate the selection institution.[12] Thus, if men and women exhibit the same behavior in the volunteer conditions but their behavior differs in the electoral conditions, then we must attribute those differences to something other than abilities or confidence.

Another avenue through which the math task might induce gender differences in behavior is not through beliefs in the decision calculus but instead by depressing women's performance on the task itself. This distinct phenomenon is referred to as *stereotype threat.* As Spencer, Steele, and Quinn (1999) define it, stereotype

---

[11] See the Theoretical Appendix in the supporting information for a comprehensive formal analysis. Note that while we state the decision rules as cutoff rules, it is straightforward to extend the analysis to a random utility framework. Doing so implies that the probability of volunteering is increasing in s_i and decreasing in $E[v_j]$, which has direct implications for a probit model.

[12] Throughout our discussion, it should be understood that the usual "ceteris paribus" assumption applies.

threat occurs when "one's performance in situations where that ability can be judged comes under an extra pressure—that of possibly being judged by or *self-fulfilling* the stereotype—and this extra pressure may *interfere with performance*" (6, emphasis added). We believe this unlikely because, as Spencer, Steele, and Quinn (1999) demonstrate, stereotype threat manifests in difficult tasks (e.g., GRE exams) rather than easy tasks (e.g., simple arithmetic problems, like ours) and can be eliminated when the task is presented as gender neutral (which we do). Even so, because stereotype threat operates through task performance independent of the institutional environment, we should see the effects of stereotype threat in *both* electoral (candidate) and nonelectoral (volunteer) settings. And we cannot overemphasize the fact that math-induced stereotype threat is a factor that can *only* operate by depressing task performance. If gender differences vary across the selection mechanisms, then we would be able to eliminate stereotype threat as a potential confounding factor in our study.

Candidates: Election Aversion

If candidate entry decisions hinge on factors related specifically to electoral competition—such as inhibitions about asking for votes or a lack of trust in the electoral system—rather than on abilities or confidence in one's relative ability as a representative, then we would expect to see gender differences arise not in the volunteering decisions (Part 2) but in the candidate entry decisions (Part 3). In other words, because Parts 2 and 3 differ only in the selection mechanism and are otherwise identical in terms of the task and payoffs, if we observe that men and women volunteer at equal rates but that women choose to become candidates less frequently than men, then we can uniquely attribute the difference in behavior to what we call *election aversion.*

We posit two specific sets of beliefs related to election aversion.[13] The first factor is an individual's *beliefs about electoral competition.* Individuals are more likely to run if they expect fewer other group members to run. Therefore, if men and women hold different beliefs about the degree of competition (i.e., if men underestimate or women overestimate it), then we would expect to see gender differences in candidate entry decisions but not in volunteering decisions, even holding ability constant.

The second factor is an individual's *beliefs about the informativeness of campaigns.* Beliefs about the informativeness of campaigns depend, in turn, on the degree of

honesty one expects of other group members as well as one's own aptitude for conveying to others one's own task ability. Gender differences in entry might therefore arise if men and women hold different beliefs about the nature of elections. We would expect women to be less likely than men to run if they believed elections were less informative—in other words, if they tend to believe (more than men do) that elections are less about merit and more about strategic posturing, misrepresentation, or vote seeking.

We experimentally manipulated the informativeness of the selection mechanism in order to test for gender differences in election aversion and to assess whether such differences arise from beliefs about elections. On the one hand, the *Truthful* campaigns are guaranteed to be perfectly informative. There is no concern that candidates will misrepresent their scores or that a candidate's report about her score will not be believed. The *TNO* and *TCB* conditions therefore represent situations in which all group members must believe that elections are perfectly informative, as differences in such beliefs cannot explain differences in behavior. On the other hand, if one believes an election to be a completely random event in which every candidate is equally likely to win, then such an election would be identical to the random selection mechanism in the *Volunteer* conditions.

Differences between men's and women's expectations about the nature of elections will have the greatest effect on decisions in the *Chat* environment, as the informativeness of the campaign in the *CNO* and *CCB* conditions arises endogenously from group behavior. Thus, we can draw inferences about whether gender differences arise from such beliefs by comparing behavior in the *Chat* conditions to the behavior in the other conditions. If men and women alike believe that campaigns tend to be truthful, then there would be no differences in behavior between the *Chat* and *Truthful* conditions. Similarly, if men and women alike believe that campaigns and elections are completely random, then there would be no differences in behavior between the *Chat* and *Volunteer* conditions. But if women tend to be more pessimistic about elections or their electoral prospects (e.g., fearing their campaigns to be less effective than others' or that other candidates will distort their messages substantially), then we expect to observe fewer female candidates in the *Chat* conditions than in either the *Truthful* or *Volunteer* conditions.

Our experimental manipulations therefore allow us to test for the existence of election aversion and to identify its potential sources by examining the patterns of gender differences produced by our alternative institutions. If we observe gender differences in the *Chat* conditions but not in the *Volunteer* decisions, then we will be confident

---

[13]See the Theoretical Appendix in the supporting information for a precise formalization and comprehensive analysis.

that elections themselves are a cause of the gender gap in candidate decisions. If this gap shrinks or disappears when campaigns are *Truthful*, then we can attribute election aversion to beliefs about the informativeness of the election. Otherwise, our theoretical analysis implies that different beliefs about the competitiveness of elections might be to blame, and this can be tested directly using data from our belief elicitation task.
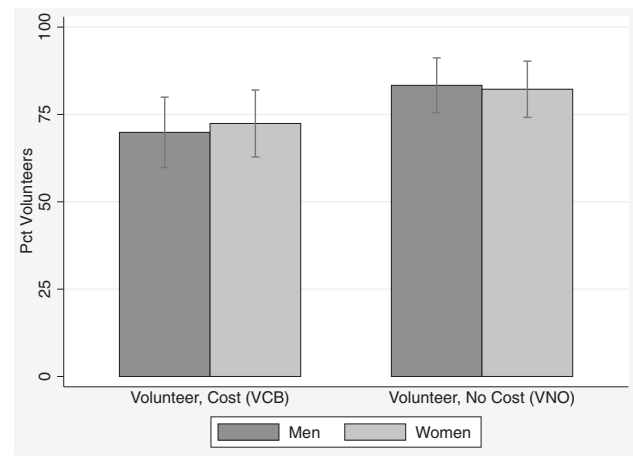
Risk Aversion

The above arguments seem to depend on the assumption that men and women are both risk neutral, but prior research suggests that women tend to be more *risk averse* than men (Croson and Gneezy 2009; Eckel and Grossman 2002). Risky decisions are those for which there is greater uncertainty—more specifically, greater variance—in the outcomes. Individuals who are risk averse prefer to avoid or minimize the amount of such uncertainty. To determine how risk aversion affects the decision to run for office, we must therefore consider whether it is more or less risky to be a candidate. Note that the risky outcome in question is the representative's score, and so an individual's decision rule must take into account whether being a volunteer or candidate increases or decreases the variance of this outcome.

In general, there is *less* risk in being willing to be the representative than in not being so willing. The reason is that an individual is much more certain about his or her own score than the scores of others. By volunteering or running for office, an individual increases the probability that her own score will be the representative's score, thereby reducing the variability of the overall outcome. Formally, risk-aversion lowers the threshold for entry. Stated more intuitively, risk averse-subjects prefer to volunteer because there is a greater chance that the representative's score will be one they know (their own) than one they do not (another group member's score). This argument also holds if subjects view their own scores as variable rather than fixed.[14]

FIGURE 3 Volunteer Choices

If women are generally more risk averse than men, then they will be *more likely* to volunteer than men with the same ability and beliefs, especially in the *VNO* condition. Of course, this prediction also depends on whether subjects correctly recognize that their entry decisions reduce uncertainty. Nevertheless, we emphasize that if differential risk aversion contributes to gender differences in entry decisions, we would expect to observe such differences across *all* conditions of our experiment. That is, differences in risk aversion would not explain differences between the volunteering or candidate entry decisions or between candidate decisions when there are differences in the campaign environment.[15]

Results
Who Volunteers? Who Runs?

Our results demonstrate that men and women make the same choices when faced with the decision to volunteer but reveal dramatic differences in how men and women approach the question when the selection of a representative requires an election. Figures 3 and 4 present the main results for subjects' willingness to be the group representative for each of our selection mechanisms. As shown

---

[14]More formally, suppose that an individual's performance on the task is a function of underlying ability a_i and some random noise ε_i; that is, $s_i = a_i + \varepsilon_i$. Because others' scores are unknown, a_j for $j \neq i$ is a random variable, and the variance of some other group member's score s_j is $Var(s_j) = Var(a_j) + Var(\varepsilon_j)$. However, because a_i is known, the variance of one's own score is $Var(s_i) = Var(\varepsilon_i)$, which will be lower than $Var(s_j)$ if $Var(\varepsilon_j) \geq Var(\varepsilon_i)$. Even if an individual thinks her own score is more variable than others', it is still likely that the variance of a_i is lower than the variance of a_j; in order for volunteering to be the riskier decision, one's own performance variance $Var(\varepsilon_i)$ must exceed the *combined variance* of others' scores $Var(a_j)$ and their performance variance $Var(\varepsilon_j)$.

[15]We can also test this explanation directly using our elicited measure of risk preferences, which provides a measure of the curvature (local concavity) of subjects' utility for money. While our conception of risk *preference* stems directly from expected utility theory, even if we adopted a psychological conception of risk *attitudes* (e.g., Weber, Blais, and Betz 2002), such a theory would also predict the same patterns of behavior across institutions. That is, if gender differences in psychological risk attitudes account for the gender gap in candidate entry, then we would observe this gap across both volunteering and candidate entry decisions.

FIGURE 4 Candidate Entry Choices
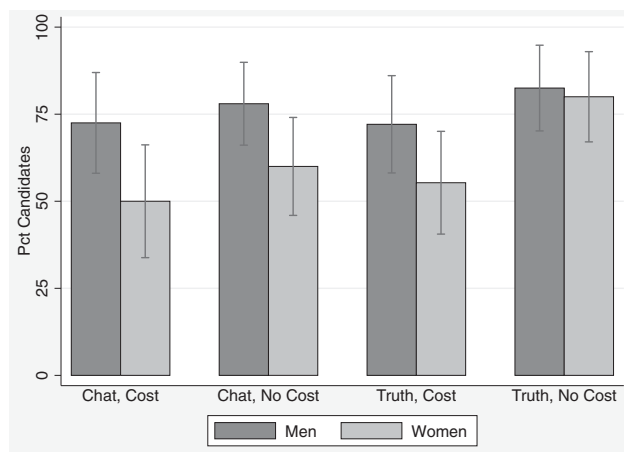
in Figure 3, when selection is random, both men and women are very likely to volunteer in the baseline *VNO* condition (Part 2 without private costs and benefits), and they do so at similar rates (83.3% and 82.2%, respectively, $\chi^2_{(1)} = 0.04$, p = .84). Adding costs and benefits decreases the rates of volunteering (to 69.9% for men and 72.4% for women), but the rates remain similar and statistically indistinguishable ($\chi^2_{(1)} = 0.13$, p = .72). These results indicate that we have successfully controlled for gender-based differences in ambition. Because the task is identical in the volunteer and election treatment, any desire to hold office would affect the decision in both stages equally. Given the fact that we do not see gender-based differences in the volunteer decision, if either relative confidence in ability or level of ambition is the major factor determining whether or not subjects are willing to represent their group, we would also not expect to see differences in the candidate decisions.

Yet this is not the case. Figure 4 shows that there are substantial gender differences in the willingness to enter the pool of potential representatives when the selection mechanism involves an election. In the *CCB* condition, which reflects two important features of electoral processes we believe operate in real-world environments (campaigns and private costs and benefits), the percentage of male candidates (72.5%) remains similar to the percentage of male volunteers in the *VCB* condition. But the percentage of women who run drops substantially, to 50%, and this difference between men and women is statistically significant ($\chi^2_{(1)} = 4.27$, p = .04).

The results also show that when we experimentally manipulate only one of the electoral factors, women continue to exhibit a high degree of election aversion, whereas men choose to become candidates at the same rates that
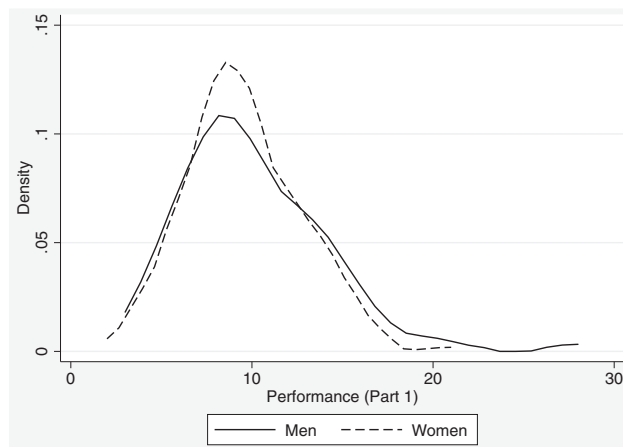
they chose to volunteer. In the *CNO* condition, when we remove costs from the electoral environment while retaining chat campaigns, 60% of women run for office compared to 78% of men, and this gender gap is also statistically significant ($\chi^2_{(1)} = 3.79$, p = .05). Similarly, in the *TCB* condition, when we retain costs but make campaigns truthful, 55.3% of women run, whereas 72.1% of men do (however, this difference does not reach conventional levels of statistical significance, $\chi^2_{(1)} = 2.72$, p = .10).

The exception to this pattern occurs when we simultaneously guarantee that elections truthfully reveal candidates' abilities and remove the direct costs of entry. In the *TNO* treatment, 80% of women run for office, a rate that is statistically indistinguishable from that of men (82.5%; $\chi^2_{(1)} = 0.08$, p = .78) and comparable to the rate of volunteering in the *VNO* condition.[16] The prospect of campaigning and of bearing the costs of participating in the electoral process each appears to be individually sufficient to cause greater election aversion in women while having no effect on men.

Our findings point to three broad conclusions about gender and election aversion as a distinct behavioral phenomenon. First, our results demonstrate that men and women differ dramatically in their willingness to stand for election, even in the absence of external forces that the extant literature suggests are important, such as family obligations, access to money, or political socialization. Second, and more significantly, because we find that women and men volunteer at similar rates, neither the relative lack of confidence in ability nor differences in risk aversion appear to explain these gender differences, which instead appear to have uniquely electoral causes. Third, our findings suggest that women's decisions are more sensitive to the institutional context: Eliminating differences in election aversion requires removing the strategic elements of campaigning (i.e., increasing the informativeness of the election) as well as removing (or reducing) the

---

[16]The fact that our manipulation of the electoral institution eliminates the gender difference constitutes evidence against several alternative explanations. As we explained in the theoretical analysis, even if we do not have good measures of risk preferences or attitudes, the results of the *TNO* treatment effectively rule out differences in risk aversion as the source of the gender difference. This also holds for unobserved differences in beliefs about the variability of task performance, expectations of future performance, or altruism and other-regarding preferences. It also rules out the possibility that women are more likely to engage in "volunteer rolloff" due to the order of the nonelectoral and electoral treatments. That is, some women may decide not to run in Part 3 because they had already volunteered in Part 2. Although this is unlikely because we randomly select one part for payment, making the decisions in Parts 2 and 3 independent decisions, the elimination of the gender difference in the *TNO* treatment provides evidence against this interpretation.

FIGURE 5 Task Performance (Kernel Densities)

Performance (Part 1)

TABLE 1 Beliefs

	Estimates			
	Men	Women	p-Value	Actual
All Treatments				
Highest score	12.7	12.8	.96	13.7
2nd highest score	10.5	10.6	.97	10.7
3rd highest score	8.7	8.6	.63	8.6
Lowest score	7.0	6.7	.29	6.4
Average score	9.7	9.7	.74	
Volunteer, Cost				
Number of volunteers	2.2	2.4	.18	2.8
Average volunteer score	10.0	10.2	.83	10.5
Volunteer, No Cost				
Number of volunteers	2.6	2.6	.73	3.3
Average volunteer score	10.5	10.3	.81	10.4
Election, Chat + Cost				
Number of candidates	2.0	1.8	.32	2.5
Average candidate score	9.3	9.8	.62	10.6
Election, Chat + No Cost				
Number of candidates	2.3	2.2	.61	2.8
Average candidate score	11.0	9.6	.09	10.5
Election, Truth + Cost				
Number of candidates	1.9	2.0	.37	2.5
Average candidate score	10.0	10.8	.41	10.5
Election, Truth + No Cost				
Number of candidates	2.3	2.6	.21	3.3
Average candidate score	9.7	11.0	.14	10.2

Note: The p-values are for two-tailed difference-in-means tests.

financial downsides to running for office (i.e., decreasing the relevance of expected competition).

To determine more carefully whether election aversion may be distinct from other (potentially related) gender-based factors, we next analyze the data generated by the addition task itself as well as from our incentivized belief and risk measurement tasks.

Abilities

We find no differences in the distribution of task performance by gender. Aggregating across all experimental conditions, the mean number of correct sums in Part 1 is 10.1 for men and 9.5 for women. This difference is not statistically significant (p = .14, two-tailed). Although there is a slight difference in the dispersion of the distributions (the standard deviation is 4.1 for men and 3.1 for women) that appears to be due to slightly heavier right tail for men (shown in Figure 5), a Kolmogorov-Smirnov test fails to reject the equality of distributions (p = .66). Thus, we find that men and women are equally qualified to be the group representative. Moreover, the fact that the distribution of scores is equal also rules out the possibility that stereotype threat is a relevant factor in our experiment. If it were, we would see women underperforming relative to men due to the extra psychological pressure of feeling judged by a negative, self-fulfilling stereotype. This does not appear to be the case.[17]

Beliefs

The results of our belief elicitation task demonstrate that men and women hold nearly identical beliefs about the abilities of others. As the first four rows of Table 1 show, the mean guesses of the highest other group members are 12.7 for men and 12.8 for women; for the second highest are 10.5 and 10.6; for the third highest are 8.7 and 8.6; for the lowest are 7.0 and 6.7. The implied average belief is 9.7 for both men and women.[18] None of these slight differences are statistically significant. Not only are mean beliefs by gender nearly identical, but they are also quite accurate (compare the guesses to the actual scores shown in the right-hand column). Our experimental evidence therefore suggests that differences in beliefs about relative ability (e.g., that might arise from negative stereotypes

[17]In a post-experiment questionnaire, we also asked subjects to write what they thought the experiment was about. Only two subjects mentioned race but thought the experiment was about performance rather than elections. None of the subjects mentioned gender. These results suggest that we successfully avoided priming or cuing gender considerations.
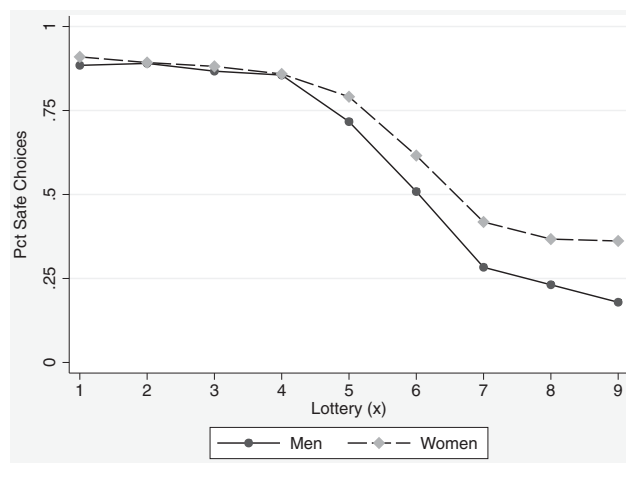
[18]Although subjects do not guess the averages directly, we can compute them from the set of beliefs we elicited.

FIGURE 6 Lottery Choices

about math abilities) do not explain the gender differences between candidate entry that we observe.

Turning next to beliefs about competitiveness, men and women also form indistinguishable beliefs about the decisions of others to join the pool of candidates. Although subjects of both genders tend to underestimate the number of volunteers and candidates, there are no statistically significant differences between their estimates. There are no differences in the implied average scores either. In general, we also find that beliefs appear to be responsive to the manipulation of the selection mechanism. For example, the mean number of expected volunteers or candidates and their average scores are higher in the treatments without costs and benefits than in the treatments with costs and benefits, as payoff maximization would imply. This suggests that subjects recognize the role that costs and benefits play in the willingness to represent one's group and that they recognize that other subjects' actions will depend on them as well. Gender differences in election aversion do not appear to be a matter of men and women perceiving their competition differently.

Risk Preferences

We find that in our risk elicitation task, women tend to choose the safer option more often than men. Recall that we designed our second incentivized measurement task so that subjects faced a series of gambles that capture the monetary risk involved with the decision to volunteer. Figure 6 shows the percentage of men and women who chose the safer (less risky) option for each of the nine binary choices. Consistent with expected payoff maximization, we find that as the expected value of the risky option

increases relative to the safer option (as x increases), the proportion choosing the safer option decreases, and this pattern holds for both men and women. The fact that 72% of men and 79% of women choose the safer option when A and B are equal in expected value (when $x = 5$) suggests that both genders exhibit some degree of risk aversion.

When there is a trade-off between the higher expected value of the risky option and the lower risk of the safer option (the lotteries where $x > 5$), we see that women are more likely to choose the safer option than men; all of these gender differences are statistically significant. When the expected benefit of choosing the risky option is greatest (for $x = 9$), the proportion of men who chose the safer gamble drops to 18%, whereas the proportion of women who made the same choice is twice as high, at 36%. On average, the total number of safe choices is higher for women (6.1) than for men (5.4), and this difference is statistically significant. Thus, consistent with previous research on behavioral measures of risk preference, we find that women exhibit greater risk aversion than men.

However, the fact that men and women differ in their risk preferences does not, in and of itself, implicate differential risk aversion as responsible for differences in candidate entry decisions. Given that men and women have the same abilities and the same beliefs about others, differences in risk preferences should lead women to volunteer less than men in the *VNO* and *VCB* conditions, especially given that the randomness of the selection process helps to focus attention on the element of risk. But this is not what we found. Because we see gender differences only in the election condition, this casts doubt on the explanation that risk aversion accounts for differences in candidate entry.[19]

Probit Analysis

We next estimate a series of probit regression models to more carefully assess how changes in the selection mechanism, task ability, beliefs, and risk preferences independently affect the decision to enter the pool of potential representatives.[20] We estimate the models separately for men and women so that we can investigate whether these

[19]In additional analysis (reported in the supporting information), we find that a survey measure of risk attitudes based on Kam (2012) is not correlated with election aversion.

[20]In the analysis, we stack the data making the dependent variable the *Choice* to represent so that there are two observations for each subject (one for the choice to volunteer in Part 2 and one for the choice to be a candidate in Part 3). This allows us to estimate both the between-subjects and within-subjects effects within the same statistical model. The scores are mean-adjusted so that the average

TABLE 2 Probit Analysis of Volunteer and Candidate Entry Choices

	Men			Women		
Score	0.16*	0.36*	0.31*	0.12*	0.25*	0.38*
	(0.03)	(0.05)	(0.07)	(0.03)	(0.05)	(0.09)
Volunteer Cost	−0.39	−0.24	−0.15	−0.39	−0.39	−0.61*
	(0.23)	(0.25)	(0.26)	(0.21)	(0.23)	(0.30)
Election Chat-Cost	−0.21	−0.04	−0.05	−1.03*	−0.90*	−1.10*
	(0.28)	(0.31)	(0.32)	(0.25)	(0.27)	(0.33)
Election Chat-No Cost	−0.33	−0.16	−0.12	−0.71*	−0.57*	−0.80*
	(0.27)	(0.30)	(0.29)	(0.21)	(0.25)	(0.29)
Election Truth-Cost	−0.39	−0.11	0.07	−0.85*	−0.64*	−0.86*
	(0.27)	(0.29)	(0.33)	(0.25)	(0.27)	(0.32)
Election Truth-No Cost	0.08	0.26	0.32	−0.16	−0.10	−0.34
	(0.28)	(0.28)	(0.35)	(0.26)	(0.29)	(0.32)
Safe Choices		−0.07	−0.06		−0.08	−0.09
		(0.05)	(0.05)		(0.05)	(0.05)
Believed Number Others		0.34*	0.34*		0.51*	0.53*
		(0.09)	(0.08)		(0.08)	(0.08)
Believed Average Score		−0.26*	−0.26*		−0.18*	−0.18*
		(0.05)	(0.05)		(0.04)	(0.04)
Volunteer Cost × Score			0.08			−0.15
			(0.08)			(0.11)
Elect Chat-Cost × Score			0.02			−0.03
			(0.09)			(0.11)
Elect Chat-No Cost × Score			−0.03			−0.17
			(0.07)			(0.10)
Elect Truth-Cost × Score			0.19			−0.16
			(0.12)			(0.11)
Elect Truth-No Cost × Score			0.06			−0.27**
			(0.12)			(0.13)
Constant	1.04*	3.48*	3.35*	1.05*	2.41*	2.71*
	(0.18)	(0.69)	(0.70)	(0.15)	(0.66)	(0.66)
Log likelihood	−162.01	−135.9	−133.57	−193.85	−157.71	−154.25
N	346	346	346	354	354	354

Note: Each column reports coefficient estimates for a separate probit regression model. Robust standard errors clustered by subject in parentheses. *p < .05.

factors affect their decisions differently. The first model, reported in Table 2, includes task ability and an indicator variable for each selection condition (*VCB, CNO,* etc.), with *VNO* as the omitted category. The results show that while men and women are equally sensitive to task performance and equally responsive to the addition of costs

score variable is 0; we do this to make the treatment coefficients more interpretable when we include interactions in the model (as the shift in the probability of entry by an average member). To correct for within-subject dependence, we cluster the standard errors by individual. See the supporting information for complete details and additional analysis showing that the results reported here are robust to alternative specifications.

in the *VCB* condition, women are sensitive to the electoral context, whereas men are not. Even when controlling for task ability, women are less likely to run in all of the election treatments except for the *TNO* condition.

In the second model, we add the elicited measures of beliefs about the number of other volunteers/candidates, their implied average scores, and a measure of risk aversion. Beliefs affect the decisions of both men and women in similar ways, whereas risk aversion is not significant. But adding these measures to the statistical model leads only to a small decrease in the estimated coefficients for the election effects for women and does not eliminate them, meaning that beliefs and risk aversion cannot

account for women's responsiveness to electoral considerations. This responsiveness remains in a third specification in which we add interaction terms that allow subjects' responsiveness to task ability to vary across selection mechanisms. Overall, our analysis demonstrates that election aversion among women seems to be a robust behavioral phenomenon that cannot be explained by differences in ability, differences in beliefs, or differences in risk aversion.

Social Preferences

Up to this point, we have interpreted our experimental results in light of a seemingly narrow theory of individual payoff maximization. But previous research suggests that men and women differ in their level of prosociality (Andreoni and Vesterlund 2001; Eckel and Grossman 1998), finding also that such differences can be context dependent (Croson and Gneezy 2009). Would expanding our theoretical framework to include other-regarding or *social preferences* therefore help to explain why women's entry decisions are responsive to the institutional context and to electoral factors in particular?

In this section, we analyze several specific forms of social preferences. We argue that social preferences such as altruism, inequality aversion, and "warm glow" motivations cannot explain our findings. As we explain, the payoff structure and institutional manipulations in our experiment effectively rule them out. However, it is possible that differences in honesty or trust might explain differences in candidate entry decisions.[21] It is important to note that these latter factors do not constitute alternative explanations, as they operate specifically in the electoral context, but rather provide potential mechanisms for election aversion.

Altruism

In the *VNO, CNO,* and *TNO* conditions, any potential gender differences in altruism or inequality aversion are *irrelevant* to the decision to volunteer or run for office when there are no private costs or benefits to the representative. Indeed, we designed the incentives in the experiment such that individual and group incentives are *completely aligned.* Doing what is best for oneself is

exactly the same as doing what is best for the group.[22] If an above average individual volunteers, she raises the expected value of the representative's score, which increases her own expected payoff. But note that doing so raises the expected payoff of all fellow group members as well. Similarly, if a below average individual volunteers, she lowers her own expected payoff but also harms the group by lowering the expected payoffs of every group member. Thus, when there are no private costs or benefits, differential altruism cannot explain differences in behavior because selfish and altruistic individuals would do the same thing.

The argument against inequality aversion is similar. If an individual is inequality averse, she has a preference for outcomes in which every group member's payoff is more similar than dissimilar. That is, she simply prefers outcomes in which there is less inequality. But the decision to volunteer has no effect on the inequality of outcomes. This is because increasing or decreasing the representative's score has an equal effect on every member's payoff. Because altruism and inequality aversion are irrelevant in the no-cost conditions, the differences we observe between the *VNO, CNO,* and *TNO* conditions cannot be attributed to differences in social preferences between men and women.

Once we introduce private costs and benefits in the *VCB, CCB,* and *TCB* conditions, altruistic preferences begin to have some theoretical relevance, but they do not provide a satisfactory explanation for women's behavior. First, the effects of altruism should be relatively small in magnitude because they are limited to how an individual weighs the costs and benefits of office. And they occur only at the margin, as volunteer and candidate entry decisions will still largely be determined by whether one has a relatively high or low score. Second, given a sufficient level of competition, greater altruism would predict a higher probability of becoming a volunteer or candidate.[23] Moreover, differences in altruism would also have

---

[21] We do not consider reciprocity. Unlike the trust game, ultimatum game, or repeated social dilemmas, the group representation decision in our experiment provides no opportunities for cooperation or reciprocal altruism.

[22] To specify formally how other-regarding preferences such as altruism would affect the decision, let $\alpha \in [0, 1]$ denote the weight that subject i places on the payoffs of the other members of her group, where $\alpha = 0$ indicates that i is completely selfish and $\alpha = 1$ indicates that she considers the payoff of each group member to have equal weight as her own. In the *VNO* condition, the decision rule is identical for all possible values of α: Volunteer if and only if one's score is above average. Similarly, the decision rules in the *CNO* and *TNO* conditions will also be independent of α.

[23] Formally, if α represents the relative weight that a subject places on others' payoffs, the decision rule is to volunteer if $s_i \geq E[v_j] + \frac{4\alpha + 2n - 2}{1 + 6\alpha}$, where n is the belief about the number of other group members who have volunteered. Here, the effect of altruism (increasing α) depends on the number of other volunteers expected. When $n \leq 1$, then the more other-regarding a subject is, the less likely she is to volunteer (allowing someone else to obtain the benefit of being selected). But when $n > 1$, she is more likely to

the same effects on the volunteer decision as on the candidate entry decisions. The observation that women are less likely to become candidates (holding beliefs constant) coupled with the observation that women make *different* decisions in the *VCB* condition than they do in the *CCB* or *TCB* conditions suggests that their behavior cannot be attributed to greater other-regarding considerations.[24]

Another form of altruism is the "warm glow" theory (Andreoni 1990). Instead of caring directly about others' payoffs, an individual might receive personal satisfaction from choosing an action that is perceived as benefiting others. In our experiment, subjects might receive an unobserved psychological benefit (rather than a monetary benefit) from serving others as the group representative. Individuals motivated by such "warm glow" are therefore more likely to be candidates or volunteers than others. Note that being the representative in the volunteer and the election conditions entails an identical task and therefore involves serving others in exactly the same way. If women are motivated by a stronger sense of "warm glow" altruism than men, then we would expect their behavior to be the same as volunteers and as candidates: They would be more likely to volunteer *and* be more likely to become candidates. But this is not what we observe (as we just noted with respect to pure altruism), thus ruling out "warm glow" as an explanation of women's behavior.

Trust and Honesty

The kinds of social preferences that might predict different behavior in electoral environments pertain to trust and honesty. Although trust is a broad concept with many dimensions (e.g., Glaeser, Laibson, Scheinkman, and Soutter 2000; Wilson and Eckel 2011), the particular form that may be relevant in our experiment is the expectation that others will portray themselves honestly. In other words, men and women might have different levels of trust that other candidates will engage in truthful campaign behavior. This form of trust is very similar to what we mean by the informativeness of campaigns. If women are less likely to "trust" that others will be honest, then they will believe campaigns to be less informative and, as we explained in our theoretical analysis, we would

volunteer because the social cost of volunteering is small relative to the social benefit of increasing everyone's payoffs.

[24]As an additional check to see whether altruistic or prosocial attitudes might be correlated with behavior, we included several personality batteries with additional sessions of the *CCB* treatment and found that none of the resulting personality scales mediate election aversion among women. See the supporting information for details of this analysis.

observe that women are less likely to become candidates when there is a *Chat* campaign than a *Truthful* campaign. (Of course, such trust plays no role in the decision to volunteer, so differences in trust cannot explain volunteering decisions, only candidate decisions.) Although we do not have direct evidence for subjects' trust about campaigns, we do find that increasing the truthfulness of campaigns (in the absence of costs) increases the entry of women candidates, which is consistent with how trust might influence decisions.

An additional consideration is the degree to which one is willing to lie (Gneezy 2005) or, relatedly, the degree to which one expects lies to be believed. Suppose, for example, that candidates believe that campaigns tend to be dishonest and that candidates can win by inflating their scores. Even if a potential candidate is confident that she is among the highest-ability performers and therefore believes it is in both her own and her group's best interest to be the representative, she may be unwilling to do so if she is *lying averse*—that is, if she bears high psychological costs for lying or if she places a high value on acting honestly. If women are more averse to lying than men (e.g., Dreber and Johannesson 2008), then they would be more likely than men to avoid running for office in order to avoid incurring the costs of lying. Note, however, that such differences can arise only in the *Chat* campaigns because *Truthful* campaigns are honest by design and the *Volunteer* decision involves no campaign; any effect of lying aversion must be specific to the campaign context. Indeed, our finding that women are less likely to run for office in the *Chat* environment than in the *Truthful* campaign environment is consistent with a lying aversion explanation, which suggests a potential social mechanism that might help explain election aversion.

Campaign Messages

The data hint at subtle gender differences in campaign styles, which provides suggestive evidence for the relevance of lying aversion. As we do not have messages from non-candidates, our data for campaign messages clearly suffer from selection bias, so we caution that we cannot draw any firm conclusions. Nevertheless, we performed a simple content analysis of campaign messages, identifying messages that contained specific numerical claims about performance (e.g., "I got 21 and 19 correct in the first two rounds") and messages containing vague, but relevant, claims about mathematical ability (e.g., "I'm great at math!!!!!!!!!!!"). We also checked each numeric claim against the candidate's actual scores in Parts 1 and 2 of the experiment to assess its truthfulness and coded whether

TABLE 3 **Campaign Message Characteristics**

	Chat CB		Chat NO	
	Men	**Women**	**Men**	**Women**
Vague Messages	26%	20%	13%	30%
Numeric Messages	63%	65%	77%	57%
Exaggerated Messages	20%	10%	10%	4%
Average Exaggeration	6.0	4.5	2.0	1.0
N	27	20	39	30

messages exaggerated a candidate's true performance and by how much.

Table 3 presents a simple tabulation of these results by gender and treatment. While men and women write messages containing numeric claims at similar rates in the *CCB* treatment, men have a somewhat greater tendency to exaggerate their numerical messages and to exaggerate their scores by a greater amount. Interestingly, from an individual payoff-maximizing perspective, men tend to exaggerate "too much", whereas women tend to exaggerate "just enough" (relative to the possibility of losing the election to a higher-ability group member, thereby obtaining a higher overall payoff from the representative's performance). In the *CNO* treatment, when there is no benefit to lying, the extent and degree of exaggeration decreases for both men and women. However, men still lie somewhat more often than women and exaggerate by a greater amount. The proportion of numeric messages also increases slightly for men and decreases slightly for women. Thus, even when there is no material benefit to lying, men still seem to lie too much.

The data also suggest that exaggerating works. In the *CCB* treatment, when candidates are prone to lying, the best candidate is elected in only 56% of the groups. In contrast, the best candidate is elected in 75% of groups in the *CNO* treatment. When candidates are precluded from lying in the truthful campaign treatments, the best candidate is elected in 94% of the groups in the *TCB* treatment and 100% of groups in the *TNO* treatment. We speculate that the noisiness of the campaign environment is, to a large extent, an important source of election aversion.

Conclusion

According to traditional democratic theory, elections are meant to be the catalyst for proper representation. They are what encourages representatives to translate their constituents' interests into policy outcomes. Indeed, despite their ignorance, voters seem to do reasonably well in selecting representatives (Lau and Redlawsk 1997). But if voters do not face the full range of choices because the pool of candidates is limited, then democratic representation may not be able to reach its full potential. Diverse legislatures offer many potential benefits, but legislatures cannot be diverse if the pool of candidates is not diverse to begin with.

Focusing on gender, we find that women are less likely than men to run for office in a controlled, incentivized laboratory experiment where group members have common incentives to select the best representative. This gap cannot be explained by differences in objective task abilities, subjective beliefs about the abilities of others, or risk preferences. By carefully manipulating the details of the selection mechanism while holding other features of the environment constant, we can confidently attribute the gender gap in our experiment to the particular context of campaigns and the costs of elections. Indeed, because the job of the representative is the same in all conditions of the experiment, we can infer that women are sensitive to the details of the selection process, whereas men are not.

We suggest that election aversion may be an important behavioral source of women's underrepresentation that is distinct from, but also complements, existing attitudinal and preference-based explanations. Indeed, our results are consistent with a growing body of literature in the behavioral sciences that seeks to explain that the dearth of women in a variety of important positions in society—in politics as well as in business leadership, science, and technology—is the result of not only external (e.g., Bowles, Babcock, and Lai 2007), but also *internal* motivational factors (e.g., Babcock and Laschever 2003; Gneezy, Niederle, and Rustichini 2003; Jones and Linardi 2014; Reuben, Rey-Biel, Sapienza, and Zingales 2012). No doubt, the psychological mechanism and attitudinal correlates of election aversion are related to differences in decisions in other domains, but further research will be required to say exactly what these are. We suspect that election aversion may have to do with feelings of trust and honesty: a lack of trust in others to campaign honestly, in the accuracy of the electoral system to select the best representative, or an unwillingness to lie even if doing so might be necessary for groups to elect the best leaders. Election aversion may also be related to preferences for competition (Niederle and Vesterlund 2007). While the election itself does not exert greater competitive pressure on task performance (since good performance by the representative benefits everyone), campaigns and elections involve an element of political competition

and social evaluation. Perhaps women avoid such social comparisons, even if doing so is costly to the group as a whole.

It is worth acknowledging that two factors that we held constant in the experiment—the perceived masculinity of the task and the mixed gender composition of groups—may be relevant to a broader understanding of the gender gap in the real world.[25] It may very well be the case that women are deterred from running for office by the perception that politics is an activity in the masculine domain or that they are likely to have to compete against men in order to win elections. But we claim only that election aversion is a distinct contributing factor, not that it is unique. We cannot overemphasize the methodological point that experimental control is critical to making this inference. Because the addition task and mixed gender groups are held constant, these features cannot explain the differences in behavior across institutions. Nevertheless, we agree that these features of politics might contribute to the gender gap among actual candidates in political settings and may be necessary conditions for election aversion. But importantly, our analysis also demonstrates that it is possible to eliminate the gender gap without making the task of the representative more feminine or by imposing single-sex competition, both of which would be impractical policy solutions. Instead, our results suggest that reforming electoral institutions with an eye toward encouraging a more diverse pool of candidates would require making truth telling more incentive compatible or varying the compensation for public service so that potential candidates decide on the basis of their policymaking abilities rather than the perks of office or the burdens of campaigning.[26] Such reforms, however, seem difficult enough and unlikely to materialize.

References

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100(401): 464–77.

Andreoni, James, and Lise Vesterlund. 2001. "Which Is the Fair Sex? Gender Differences in Altruism." *Quarterly Journal of Economics* 116(1): 293–312.

[25]We ran additional sessions of the *CCB* treatment with only female subjects to test whether removing mixed gender competition affects women's election aversion and found that it has no effect. See the "Mixed Gender Competition and Attitudes" section of the supporting information.

[26]For example, see Maestas, Fulton, Maisel, and Stone (2006) on the influence of costs and benefits on ambition and Fulton, Maestas, Maisel, and Stone (2006) for analysis of gender differences.

Anzia, Sarah F., and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55(3): 478–93.

Aronson, Elliot, and J. Merrill Carlsmith. 1990. *Methods of Research in Social Psychology*. New York: McGraw-Hill.

Babcock, Linda, and Sara Laschever. 2003. *Women Don't Ask: Negotiation and the Gender Divide*. Princeton, NJ: Princeton University Press.

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

Bowles, Hannah Riley, Linda Babcock, and Lei Lai. 2007. "Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes It Does Hurt to Ask." *Organizational Behavior and Human Decision Processes* 103(1): 84–103.

Campbell, David E., and Christina Wolbrecht. 2006. "See Jane Run: Women Politicians as Role Models for Adolescents." *Journal of Politics* 69(2): 233–47.

Conway, M. Margaret, Gertrude A. Steurnagel, and David W. Ahern. 1997. *Women and Political Participation*. Washington, DC: Congressional Quarterly Press.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448–74.

Darcy, Robert, Susan Welch, and Janet Clark. 1994. *Women, Elections, and Representation*. 2nd ed. Lincoln, NE: University of Nebraska Press.

Dreber, Anna, and Magnus Johannesson. 2008. "Gender Differences in Deception." *Economics Letters* 99(1): 197–99.

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. New York: Cambridge University Press, 41–57.

Duffy, John, and Margit Tavits. 2008. "Beliefs and Voting Decisions: A Test of the Pivotal Voter Model." *American Journal of Political Science* 52(3): 603–18.

Eckel, Catherine C., and Philip J. Grossman. 1998. "Are Women Less Selfish Than Men?: Evidence from Dictator Experiments." *Economic Journal* 108(448): 726–35.

Eckel, Catherine C., and Philip J. Grossman. 2002. "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk." *Evolution and Human Behavior* 23(4): 281–95.

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni. 2009. "Moral Bias in Large Elections: Theory and Experimental Evidence." *American Political Science Review* 103(2): 175–92.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economics Experiments." *Experimental Economics* 10: 171–78.

Fox, Richard L., and Jennifer L. Lawless. 2010. "If Only They'd Ask: Gender, Recruitment, and Political Ambition." *Journal of Politics* 72(2): 310–36.

Fox, Richard L., and Jennifer L. Lawless. 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.

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, Elisabeth R., Rebecca B. Morton, and Thomas A. Rietz. 1998. "Minority Representation in Multimember Districts." *American Political Science Review* 92(1): 127–44.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *Quarterly Journal of Economics* 115(3): 811–46.

Gneezy, Uri. 2005. "Deception: The Role of Consequences." *American Economic Review* 95(1): 384–94.

Gneezy, Uri, Kenneth L. Leonard, and John A. List. 2009. "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." *Econometrica* 77(5): 1637–64.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118(3): 1049–74.

Grosser, Jens, and Arthur Schram. 2006. "Neighborhood Information Exchange and Voter Participation: An Experimental Study." *American Political Science Review* 100(2): 235–48.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *The American Economic Review* 92(5): 1644–55.

Huddy, Leonie, and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37(1): 119–47.

Hughes, Melanie M. 2011. "Intersectionality, Quotas, and Minority Women's Political Representation Worldwide." *American Political Science Review* 105(3): 604–20.

Johnson, Gbemende, Bruce I. Oppenheimer, and Jennifer L. Selin. 2012. "The House as a Stepping Stone to the Senate: Why Do So Few African American House Members Run?" *American Journal of Political Science* 56(2): 387–99.

Jones, Daniel, and Sera Linardi. 2014. "Wallflowers: Experimental Evidence of an Aversion to Standing Out." *Management Science* 60(7): 1757–71.

Kam, Cindy D. 2012. "Risk Attitudes and Political Participation." *American Journal of Political Science* 56(4): 817–36.

Kanthak, Kristin, and George A. Krause. 2010. "Valuing Diversity in Political Organizations: Gender and Token Minorities in the U.S. House of Representatives." *American Journal of Political Science* 54(3): 839–54.

Kanthak, Kristin, and George A. Krause. 2012. *The Diversity Paradox: Political Parties, Legislatures, and the Organizational Foundations of Representation in America.* New York: Oxford University Press.

Krook, Mona Lena. 2009. *Quotas for Women in Politics: Gender and Candidate Selection Reform Worldwide.* New York: Oxford University Press.

Landa, Dimitri, and Dominik Duell. Forthcoming. "Social Identity and the Nature of Electoral Representation." *American Journal of Political Science* .

Lau, Richard R., and David P. Redlawsk. 1997. "Voting Correctly." *American Political Science Review* 91(3): 585–98.

Lawless, Jennifer L., and Richard L. Fox. 2005. *It Takes a Candidate: Why Women Run for Office.* New York: Cambridge University Press.

Levine, David K., and Thomas R. Palfrey. 2007. "The Paradox of Voter Participation? A Laboratory Study." *American Political Science Review* 101(1): 143–58.

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 U.S. House." *American Political Science Review* 100(2): 195–208.

Matland, Richard E., and Deborah Dwight Brown. 1992. "District Magnitude's Effect on Female Representation in U.S. State Legislatures." *Legislative Studies Quarterly* 17(4): 469–92.

Matland, Richard E., and Donley T. Studlar. 1996. "The Contagion of Women Candidates in Single-Member District and Proportional Representation Electoral Systems: Canada and Norway." *Journal of Politics* 58(3): 117–40.

McDermott, Rose. 2011. "Internal and External Validity". In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. New York: Cambridge University Press, 27–40.

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

Morton, Rebecca B. 1993. "Incomplete Information and Ideological Explanations of Platform Divergence." *American Political Science Review* 87(2): 382–92.

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

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122(3): 1067–1101.

Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton, NJ: Princeton University Press.

Reuben, Ernesto, Pedro Rey-Biel, Paola Sapienza, and Luigi Zingales. 2012. "The Emergence of Male Leadership in Competitive Environments." *Journal of Economic Behavior & Organization* 83(1): 111–17.

Sanbonmatsu, Kira. 2006. *Where Women Run: Gender and Party in the American States.* Ann Arbor, MI: University of Michigan Press.

Schwindt-Bayer, Leslie A., and William Mishler. 2005. "An Integrated Model of Women's Representation." *Journal of Politics* 67(2): 407–28.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Wadsworth Cengage Learning.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35(1): 4–28.

Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69(5): 797–811.

Stone, Walter J., and L. Sandy Maisel. 2003. "The Not-So-Simple Calculus of Winning: Potential U.S. House Candidates' Nomination and General Election Chances." *Journal of Politics* 65(4): 951–77.

Volden, Craig, Alan E. Wiseman, and Dana E. Wittmer. 2013. "When Are Women More Effective Lawmakers Than Men?" *American Journal of Political Science* 57(2): 326–41.

Weber, Elke U., Ann-Renee Blais, and Nancy E. Betz. 2002. "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors." *Journal of Behavioral Decision Making* 15(4): 263–90.

Wilson, Rick K., and Catherine C. Eckel. 2011. "Trust and Social Exchange". In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York: Cambridge University Press, 243–57.

Wolbrecht, Christina. 2000. *The Politics of Women's Rights: Parties, Positions, and Change.* Princeton, NJ: Princeton University Press.

Woolley, Anita Williams, Christopeher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. "Evidence of a Collective Intelligence Factor in the Performance of Human Groups." *Science* 330(6004): 686–88.

Woon, Jonathan. 2012a. "Democratic Accountability and Retrospective Voting: A Laboratory Experiment." *American Journal of Political Science* 56(4): 913–30.

Woon, Jonathan. 2012b. "Laboratory Tests of Formal Theory and Behavioral Inference". In *Experimental Political Science: Principles and Practices*, ed. Bernhard Kittel, Wolfgang J. Luhan, and Rebecca B. Morton. New York: Palgrave Macmillan, 54–71.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

1. Measurement tasks
2. Theoretical analysis
3. Session information and sample characteristics
4. Probit analysis
5. Mixed gender competition and attitudes
6. Experiment instructions