Elections, War, and Gender: Choose to Run, Choose to Fight

Yon Soo Park

Sarah Hummel

Stephen Chaudoin

April 29, 2021*

Abstract

Most explanations of so-called "Iron Ladies" - women leaders associated with interstate conflict - emphasize gendered aspects of international politics. We highlight a different explanation based on the self selection of women into candidacy for group leadership. More competitive women are both more likely to run for office and to choose hawkish policies once elected. We demonstrate this with a laboratory experiment using online real-time, group play where participants choose to run for election, conduct a simple campaign, and represent their group if elected. We find that more competitive women select into candidacy, campaign more effectively, and then fight harder in the intergroup contests than their male counterparts. These patterns appear even though our protocol stacks the deck against gender biases by anonymizing participants and shuffling groups. Our findings emphasize the agency and preferences of Iron Ladies who choose to run and, subsequently, choose to fight harder in intergroup contests.

*We thank Dayna Judge, Dani Reiter, and Dan Smith for helpful comments. We benefited from feedback at Harvard Experimental Political Science 2020, SPSA 2020 and APSA 2019. Steffen Beverborg provided excellent programming help from SoPHIE.

1 Introduction

> *"Of course, when you're a woman and you're combative, they say you're an Iron Lady. Let me tell you" — and she snaps to like a crossbow after it has flung its arrow — "if you hadn't got a spine which was strong and firm, and a will which was strong and firm, we would never have got through."*
>
> – Margaret Thatcher, quoted by Gail Sheehy, *Vanity Fair* 1989

There is a complex relationship between gender and conflict. On the one hand, a growing body of research consistently finds that female citizens prefer more peaceful policies than male citizens. Yet, research focusing on female *leaders* often finds that female heads of state choose more hawkish policies than their male counterparts. Even a cursory look at the historical record reveals the existence of female leaders who certainly did not shy away from interstate conflict: so-called "Iron Ladies," like Margaret Thatcher or Indira Gandhi.

Why might women in leadership roles take particularly hard-line positions on issues of security, while women in the general public espouse more pacifist opinions? In other words, how do we reconcile the stereotypes of female pacifists and Iron Ladies? Existing work emphasizes the gendered nature of the strategic environment facing female leaders, where they may be perceived as weak, and compensate accordingly. We do not doubt this. However, we draw on a growing body of research on gendered aspects of domestic democratic selection, to show that Iron Ladies also self-select into democratic elections, which has a large effect on the subsequent behavior of female leaders in conflictual situations. In other words, women in leadership positions are not representative of the broader population of women, especially when it comes to conflict. While women may be more pacific on average than men, the same is not necessarily true of women who seek to and ultimately become leaders. We thus provide evidence bearing on the questions anticipated by Enloe (1989): female leaders are often outliers who - in Margaret Thatcher's words - "got through."

We focus on the key decisions made by women during and after the democratic leader selection process: whether to self-select into candidacy, how to conduct campaigns and which policies to pursue as leaders. A large body of literature documents competition aversion among women, which also manifests as an election aversion, limiting the type of women who choose to run to those who are particularly competitive.[1] The most competitive of these women are also likely to work harder on the campaign trail, making their selection, independent of gender, more likely. Once elected, they are also more likely to choose hawkish policies in competitive situations. This self-selection process explains part of why female leaders are more conflictual than their male counterparts.

Disentangling the effects of endogenous electoral selection and the strategic environment facing women is very difficult with observational data. Quantifying self-selection requires identifying a pool of potential political leaders and measuring their characteristics before they run for and win an election.[2] It is also difficult to find real-world settings where the gendered nature of strategic interactions is absent and the selection mechanism is isolated. By using a laboratory experiment, we precisely assess the effects of gendered selection on intergroup conflict behavior in an electoral and intergroup strategic environment without gender markers.

We achieve the first objective – tracing the full path from self-selection to the ultimate choices made by group leaders – with a carefully designed protocol that lets us compare how participants behave as individuals compared with when they are elected as leaders of their group. Our protocol uses online, real-time, group experiments that model conflict in a tractable way: as lottery contest games with a Tullock contest success function. As in war, players can decide to exert costly effort to win a zero-sum prize. Each player's likelihood of winning the contest equals the number of tickets she buys divided by the total number of tickets bought on both sides.[3] Each experimental session included (a) a part with one-on-one contest games in which each participant was randomly paired against another (Individual Contest Games, or ICG) and (b) intergroup contest games in

[1] See Niederle and Vesterlund (2007) on competition aversion and Kanthak and Woon (2015) and Preece and Stoddard (2015) on election aversion. Fox and Lawless (2014) find that a significant source of the gender gap in political ambition is the gap in competitive experiences.

[2] Bernhard, Shames and Teele (2020) and Folke and Rickne (2016) tackle this problem by analyzing women participants in leadership training and the career trajectories of Swedish politicians, respectively.

[3] For a recent application, see Chaudoin and Woon (2018).

which everyone was divided into two groups and group leaders were democratically elected after self-selection into candidacy and a brief campaign (Democratic Selection Games, or DSG). We estimate the effect of democratic selection on intergroup contest behavior by bench-marking against behavior exhibited by individual players in one-on-one contests.

We achieve the second objective – decreasing gendered aspects of how participants view one another – by anonymizing participants so that their gender and identity is unknown throughout. Participants can choose to be candidates, write campaign messages, and vote without knowing identifying characteristics of the other participants. Female participants are therefore "freer" to make choices without the overt influence of concerns about counter-stereotypical behavior or a fear of backlash against displaying masculine behavior.[4]

We find that female elected leaders fought much harder than their male counterparts in group contests. The difference is large. Female group leaders spent over 10% more of their group's endowment on ticket purchases than male group leaders. Our experiment thus reproduces patterns similar to observational work: female leaders are associated with more hawkish and conflictual choices made on behalf of their group.

We then show that self-selection explains the majority of these differences. We isolate and provide evidence for two mutually reinforcing mechanisms. First, women who purchased more tickets in the zero-sum contests when playing as individuals were significantly more likely to self-select into candidacy for elections. These women also placed a larger non-monetary value on winning individual contest games. Ticket purchases in the individual contests and the estimated non-monetary value of winning both capture competitiveness in non-electoral settings. Thus, we conclude that more competitive women are more likely to enter an electoral competition. These self-selection effects were much larger for women than men. Only the most competitive women chose to run for office, while the relationship between competitiveness and candidacy was much weaker for men.

Second, more competitive women were also more successful in elections, even conditioning on

[4]See Holman, Merolla and Zechmeister (2016).

self-selecting into candidacy. More competitive women worked harder within the election itself, by writing longer campaign messages. They were also more likely to choose the most successful message content, based on our categorization of different message types. Put differently, even after accounting for the effect of competitiveness on candidacy, competitiveness still affected the likelihood that a woman would become group leader. Neither of these patterns was present as strongly for men.

These findings are important because they provide evidence for an additional reason, further up the "pipeline," for why women leaders choose more hawkish policies on behalf of their groups: such choices may directly reflect their underlying preferences towards conflict and competition. As Judge (2021) puts it, this type of explanation focuses on "the interests and calculations of the female candidate herself," as opposed to only "the actors who create the structural space for a woman to attain leadership" (4). To reiterate, our goal is not to suggest that one channel, strategic environment or preferences, matters more for explaining the choices of women leaders. The two are not mutually exclusive. But our research does suggest that Iron Ladies in the real world may genuinely have more hawkish preferences, and that gendered patterns of self-selection into leadership increase the observed differences in behavior between male and female leaders.[5]

While we have contextualized our study with references to interstate conflict, many leadership decisions outside the lab and outside of international relations pertain to competitive, intergroup interactions. When party leaders negotiate over the substance of a bill or a union representative threatens a strike to resolve a disagreement over wages, they engage in zero-sum conflicts. Filibusters and strikes are akin to a costly wars of attrition, with similarities to the intergroup conflicts analyzed here. Leaders of factions within parties also compete to have their preferred party member become party leader or attain higher cabinet offices. When such leaders are democratically elected, we expect gender-based self-selection patterns to result in women pursuing more aggressive policies than men.

Finally, our study demonstrates the possibility and usefulness of a platform where participants

[5]We recognize that the concepts of sex and gender are not equivalent. We follow the literature in primarily using the term gender to describe leader attributes and their strategic environment throughout.

recruited from the broader population via MTurk can simultaneously play games, in larger groups, during a time when face-to-face interactions are limited by COVID-19. This template is extendable to a wide array of experiments that would otherwise have been conducted in an in-person laboratory setting.

2 Existing Research and Theory

Following work by Caprioli (2000) and others, a growing body of research considers how gender affects interstate conflict.[6] One strand of this research emphasizes the preferences of citizens and how these translate into a country's subsequent policy choices. Survey responses and survey experimental data consistently find that women have more pacifist preferences than men. Analyzing over 900 American public opinion survey questions covering 24 use-of-force cases from 1982 to 2013, Eichenberg (2016) demonstrates that women are less supportive of using force than men and the significance of gender persists cross-nationally. Conducting a meta-analysis of 17 recent survey experiments related to conflict, Barnhart et al. (2020) also find that female respondents are less supportive of the use of force. Other work suggest the differences across gender exist at the elite level, not just among the general polity. These studies link the representation of women in legislatures to decreases in military spending[7] and lower likelihood of involvement in Militarized Interstate Disputes.[8]

Despite this, many other studies find that female chief executives pursue especially hawkish policies. Generally, authors explain divergences between male and female leaders by highlighting the gendered strategic environment. Female politicians face additional pressure to combat stereotypes and prove their toughness and resolve. Female leaders navigate a national and international strategic environment that is biased against them. This bias is especially notable in the domain of intergroup conflict and national security. Faced with the prevailing perception that women are

[6]For recent surveys, see Reiter (2015); Sjoberg, Kadera and Thies (2018).
[7]Clayton and Zetterberg (2018).
[8]Regan and Paskeviciute (2003).

weak, female leaders may be forced to compensate by signaling their strength and resolve through more conflictual actions.

Caprioli and Boyer (2001) began with a simple analysis of conflict situations in which female leaders took part, finding they responded to crises with higher levels of violence. Several subsequent macro-level empirical studies of defense spending and the use of military force find patterns consistent with this theory. In a study of 22 democracies from 1970-2000, Koch and Fulton (2011) find that female chief executives have higher levels of defense spending and conflict, which they attribute to the pressure on female executives to combat stereotypes with more hawkish policies. Schramm and Stark (2020) show that this pattern is amplified in democracies with more constrained executives by analyzing militarized interstate disputes (MIDs) and case studies of Tansu Çiller of Turkey and Michelle Bachelet of Chile. Post and Sen (2020) extend arguments about gendered stereotyping to explain how foreign leaders respond to the hawkish postures of female executives. Studying all MIDs from 1980-2010, they find that disputes initiated by female heads of state are more likely to be reciprocated and escalate to higher levels of conflict. Powell and Mukazhanova-Powell (2019) find that the positive effect of female executives on MID initiation is most concentrated in countries with lower levels of women's rights.[9]

Micro-level studies also provide evidence of these gender effects. Swers (2007) uses structured interviews with Senate staff to reveal that female senators put more effort in building a reputation for being tough and capable in the security domain as they face additional hurdles establishing credibility on the issue. As a result, female senators were more likely to sponsor homeland security bills in the 108th Congress (2003-2004).[10] Schwartz and Blair (2020) use survey experiments to show that conciliatory policies from women are judged more harshly, creating an incentive for female leaders to take stronger positions even if, as Croco and Gartner (2014) find, survey respondents do not more harshly punish women for flip-flopping on positions.

[9]Evidence does not universally support these arguments. Horowitz, Stam and Ellis (2015) do not find a significant relationship between leader gender and conflict.

[10]On the other hand, Bendix and Jeong (2020) find that female members of the U.S. Congress were not more hawkish on foreign policy.

The Effect of Self-Selection

We contribute to this important body of research by highlighting an additional explanation for the hawkish choices of female executives: the leader selection process means that women rising to the role of chief executive are not representative of the broader population. Rather, they may be less conflict-averse to begin with and therefore choose more conflictual policies once elected. As Reiter (2015) posits in his call for future research, the "conflict attitudes of women elected to office [may not] represent the conflict attitudes of all women" (1313). Elected women may be outliers with respect to their behavior in conflictual situations. While not mutually exclusive with the presence of strategic constraints arising from gender stereotyping, the non-randomness of democratic selection is an important additional channel through which Iron Ladies emerge.

The initial step in our theory draws on Niederle and Vesterlund (2007) and a large body of subsequent literature documenting competition aversion among women.[11] In experimental economic work, numerous studies document competition aversion by analyzing whether participants choose a (competitive) tournament-based payment scheme or a (non-competitive) piece-rate scheme. Women are more likely to choose the latter, even when controlling for their objective skill and performance on similar tasks.

Competition aversion has a natural extension to elections, which are themselves a type of competition. Elections are zero-sum contests, since only one candidate can win. While most studies find there is only a limited gender bias in the election rates of comparable men and women (Schwarz and Coppock, 2020), women are less likely to run for office in the first place. This is true even if they are equally or more qualified than men. For example, Kanthak and Woon (2015) conduct a laboratory experiment in which they allow groups to choose a leader to perform an incentivized task on behalf of the group members. They find that women are less likely to be candidates for group representative when leaders are chosen via election, as opposed to random selection, and that this gap persists even when controlling for respondents' objective abilities at the task. The gender gap in candidacy only closes when they remove the cost of candidacy and

[11] For a broader survey, see Niederle and Vesterlund (2011).

reveal truthful information about past track records. Preece and Stoddard (2015) find that priming women about the competitive nature of elections decreases their likelihood of seeking additional information about candidacy, while having no significant effect on men.

Researchers generally attribute competition aversion to a variety of evolutionary and educational psychological explanations.[12] Fox and Lawless (2014) link past competitive experiences, like varsity sports or forensics, to increased political ambition. These early life experiences help explain the gender gap in political ambition. Interestingly, competition aversion and election aversion are often explained by what they are *not*. They are generally not explained by gender differences in risk tolerance or beliefs about other group members.

Competition aversion can drive gendered differences both in who runs for office *and* in the choices of elected leaders. Suppose potential candidates for leadership weigh the costs and benefits of competing for office. The benefits from entering a contest include the direct, tangible gains from winning and the personal, psychological gain you feel from winning a contest. The costs to running are also tangible, such as the time and money to be a candidate and, if a potential candidate dislikes the cutthroat competition of the electoral process, less tangible. Women who run for office must overcome this second form of disutility, which requires a greater potential benefit to compensate for the additional cost. Only the most competitive women - those for whom the potential benefits of a competition outweigh the costs - will therefore run in an election. If men do not have the same aversion to elections, then the relationship between competitiveness and their candidacy choice is not as stark.

This dynamic then influences the observed differences between male and female leaders, because there is likely a positive correlation between competitiveness when deciding whether to run for election and competitiveness in an intergroup interaction. At their core, both elections and conflictual intergroup interactions, like war, are zero sum competitions. Overcoming an internal barrier to running for office requires a degree of competitiveness that can also translate into more aggressive stances in conflict situations.

[12] Campbell (2013).

If only the most competitive women run for office, and if this correlation is weaker for men, then we expect female leaders to be more competitive relative to the general population of women, than male leaders are relative to the general population of men. If the effect is large enough, elected female leaders will be more competitive than elected male leaders, even if the general population of women is less competitive than the general population of men. Iron Ladies - that is, women who are prone to more aggressive stances on conflict - are also the type of competitive women who are most likely to run for political office in the first place. This self-selection effect is weaker for men, who do not face such strong hurdles to candidacy and, therefore, are likely to be more representative of male populations in their handling of conflict.

Additionally, we theorize that these gendered differences in competitiveness affect how female candidates campaign for office. A more competitive woman who decides to run is also likely to invest more effort into a successful campaign. Anzia and Berry (2011) find that only the most talented and hard-working women self-select into candidacy for U.S. Congressional elections. They then work harder to (co)sponsor more bills and deliver more federal spending back to their home districts, once elected. If more competitive women candidates also put more effort into campaigning, and are therefore more likely to be elected, then this could reinforce the self-selection mechanism described above. Additionally, Kanthak and Woon (2015) find that women are more sensitive to the noisiness of the campaign environment. Competitive women who opt into elections may exert greater effort on campaigning to cut through the noise. In both cases, the non-random selection of more competitive women into candidacy may be magnified by their greater success rates in the election, even after opting to run. We would therefore expect the most competitive women to not only choose candidacy, but also to work the hardest in an election and have a higher likelihood of winning the election. Since this additional effort is also a consequence of competitiveness, it should arise even though the gender of candidates is not revealed to the electorate. We formalize these expectations after describing our experimental protocol, making them specific to our empirical setup.

To be clear, our goal is to showcase the gendered effects of self-selection into democratic lead-

ership positions. Other points of the electoral pipeline are also gendered. While acknowledging their importance, we set some of these aside. For instance, there are gender disparities in party recruitment of potential candidates.[13] National security concerns make voters less supportive of female candidates.[14] In the laboratory setting, Van Vugt and Spisak (2008) and Spisak et al. (2012) both find that female leaders are less preferred in intergroup conflict situations. On the other hand, female candidates may be more likely to rise to leadership roles after national or international crises.[15] Gendered dimensions of leadership selection can extend beyond elected positions, to cabinet positions pertaining to intergroup conflict, as well.[16] Some female heads of state also gained power by birth, as in hereditary monarchies (Dube and Harish, 2020), or were aided by dynastic connections (Folke, Rickne and Smith, 2021). Our arguments are not mutually exclusive with these possibilities, but our protocol precludes us from directly evaluating their effects.

We do, however, address several other gendered selection effects within our experimental protocol. First, selection might occur based on confidence rather than competitiveness. These are related, but distinct concepts. Confidence generally refers to a belief in one's abilities to perform a task. Existing work links gendered differences in candidacy decisions to confidence gaps, where women perceive themselves as less capable than their male counterparts.[17] It is possible that gender differences in leader decisions stem from selection on confidence, as opposed to selection on competitiveness. We analyze this in later empirical sections. Our data are more consistent with selection on competitiveness than confidence, which matches findings from experimental[18] and observational work.[19]

Second, some research shows how the experience of being elected increases levels of effort in intergroup contests, an "election effect."[20] Election by one's peers heightens levels of in-group identification and increases the sense of obligation to electors, causing elected leaders to exert

[13]Fox and Lawless (2010); Ryan, Haslam and Kulich (2010); Thomas and Bodet (2013).
[14]Lawless (2004); Holman et al. (2019); Schroeder (2017).
[15]Judge (2021).
[16]Barnes and O'Brien (2018).
[17]Eg Wolak (2020).
[18]Kanthak and Woon (2015).
[19]Bernhard and de Benedictis-Kessner (2020).
[20]Citation removed for anonymity. Goldgeier (1994) finds a related effect via socialization in an autocracy.

greater effort on behalf of their group than those selected randomly. Since women often exhibit more pro-social behavior,[21] it is possible that the election effect is stronger for women. Additionally, the type of women who choose to run for office may be more attuned to and affected by these considerations than either elected men or women more generally. However, we do not find gender differences in the election effect.

3 Protocol

In observational settings it is difficult to identify a set of future leaders and measure their preferences *ex ante*. Moreover, it is impossible to go back in time to measure the underlying preferences toward competition that current leaders had before becoming elected. This makes it challenging to disentangle selection mechanisms from the gendered strategic environment women face. We use a controlled laboratory setting to focus on selection effects and how they translate to differences in observed choices of men and women leaders.

Our lab experiment has participants playing the same zero-sum contest games, first as individuals, and later in groups, which elect leaders to play against one another. This study design allows us to follow the trajectory of participants and observe their conflict choices before *and* after they are elected. We next describe the game and specify our hypotheses based on this protocol.

Game Setup

We recruited 162 participants for 10 sessions of our game from Amazon's Mechanical Turk (MTurk)in December 2019. In each session, approximately 14 participants played one-on-one and intergroup lottery contest games. In each round, participants start with an endowment of 1,000 points. They competed to win a zero-sum prize by choosing how many "contest tickets" to buy, which cost 1 point apiece. The probability of winning the prize for that round followed the Tullock contest success function, equalling the number of tickets she purchased, divided by the the sum of her

[21] Andreoni and Vesterlund (2001); Eckel and Grossman (1998).

and her opponent's ticket purchases. The players knew the value of the prize before selecting how many tickets to buy, and prize values varied by round. Participants watched an animated video that explained the game and they answered a series of quiz questions to ensure their comprehension. We encouraged participants to treat each round as a separate decision task. Their winnings were calculated based on their average earnings across 5 randomly selected rounds. Their endowments and winnings did not carry over across rounds. Participants received $1 for every 210 points, in addition to a $5 show-up fee.[22]

In each session, participants first played 12 rounds of one-versus-one contests. We call this part of each session the Individual Contest Game (ICG). We randomly paired respondents at the beginning of each round. They then played the lottery contest game against their opponent. The prize values for 12 rounds ranged from 275 to 2715.[23] Participants were not identified and the pairs were re-shuffled every round, so players did not know who they were playing against or their history. After each round, participants saw the number of tickets purchased by each player, the outcome, and their payoffs.

Participants later played 12 rounds of an intergroup contest game with democratic elections. We call this the Democratic Selection Game (DSG). We randomly placed participants into one of two groups - the "Blue Group" or the "Orange Group." Group members remained anonymous, without any identifiers. We randomly re-shuffled the groups every other round.[24]

In each round, groups chose their leaders via a democratic election. Participants first chose whether to run for election, which entailed a fixed cost of 35 points. Candidates in the election then wrote short campaign messages that were shown to their group members. The group then

[22] We have abridged the protocol description for space. For a full description of the game protocol, see Appendix A. This appendix also describes additional details about recruitment and the demographics of participants.

[23] The exact prize values across rounds are listed in Table A.1.

[24] Participants were told that there were 7 members in each group. However, it is possible that there was in fact 6 members (if there were dropouts) or 8 members. Because of the uncertainty of participant drop outs in simultaneous online intergroup games, we sometimes added an extra player. In practice, dropouts were very rare. In Appendix B, we provide a detailed rationale for this deception as well as the pre-brief and de-brief provided to the participants. Participants also played 12 rounds in groups with randomly selected group leaders, which we called the Random Selection Game (RSG). We omit analysis using these data for length. Appendix D shows a complete re-analysis of our results using data from the RSG. Our findings are consistent when benchmarking against the ICG or RSG data.

voted for their group leader.[25] The winning candidate received a bonus of 245 points, which could not be used in the contest. The election protocol thus helps us isolate two types of selection channels: (1) candidacy is not automatic, so participants must *self-select* into running and (2) there is communication between candidates and voters so candidates also select how to present themselves to voters.

The two elected group leaders then played the contest game against one another, on behalf of their group. The leaders chose how many tickets each group member would buy. That amount was subtracted from each group member's endowment. If their leader won the contest, each group member and the leader received the prize amount. In other words, the group won or lost together and leaders could not discriminate between group members in either their ticket purchases or winnings.

Note that, in terms of expected utility, the leaders' decisions in the DSG were strategically identical to those made in the ICG. Although they played on behalf of their group, from the leader's perspective, the game was identical to a 1-vs.-1 contest. Any differences in behavior between the ICG and DSG, are therefore attributable to the presence of the election. Behavior in the ICG provides data about the *ex ante* competitiveness of each participant in individual contests. This allows us to see whether certain types of participants end up selecting into candidacy and winning elections, and whether this varies across gender. We can trace the full chain from initial attributes, to electoral selection mechanisms, to observed differences between male and female leaders.

Finally, it is important to note that participant gender is never revealed to other participants. Both the election procedure and the contest games were gender-blind, removing the overtly gendered strategic incentives associated with each of these stages. We solicited participant gender, based on self-identification, during a pre-experimental survey that also collected information on other demographic attributes.

[25] Candidates could vote for themselves. Ties were broken randomly.

3.1 Hypotheses

Do female elected leaders fight harder in intergroup competitions by buying more tickets than male elected leaders? In line with observational findings that female leaders are more hawkish or conflictual than their male counterparts, Hypothesis 1 predicts that gendered selection effects are large enough that elected female leaders buy more tickets than their male counterparts in the intergroup contests.

Hypothesis 1. ***Overall Gender Difference:*** *Female elected leaders purchase more contest tickets than male elected leaders.*

Hypothesis 2 unpacks how female leaders are non-randomly selected based on their *competitiveness*. Our protocol is set up to allow *ex ante* measurements of a participant's competitiveness based on behavior in the individual contests game (ICG). We use two related measures. First, we calculate the average number of tickets the participant bought in the ICG rounds. This gives us a measure of how hard the participant was willing to fight in individual contests that has the benefit of being directly comparable to the behavior of elected leaders. Our second measure is an estimate of each participant's non-monetary value to winning (NMV). Each round presents the opportunity to win a monetary prize, but participants may also place an intrinsic value on winning. The greater this intrinsic value is, the higher the overall payoff from winning and, as a result, the more costly effort participants exert to obtain victory. One estimate of their non-monetary value to winning comes from the Nash prediction for how many tickets a participant should buy. For two, identical, risk-neutral participants, the Nash prediction for the number of tickets purchased, T, as a function of the prize value, p, and any non-monetary value to winning, v, is: $T = \frac{p+v}{4}$. Rearranging, an estimate of the NMV is therefore given by $v = 4T \quad p$. For each participant, we calculate the average value of v, based on data from the ICG rounds. We use both measures of competitiveness below. In practice, they are well-correlated with each other.

Hypothesis 2. ***Selection Effects:***

(a) More competitive women are more likely to run for election, and this relationship is stronger for women than men.

(b) Among candidates, more competitive women are more likely to win election, and this relationship is stronger for women than men.

Hypothesis 2 specifies the two channels - self-selection into candidacy and the electoral process - that contribute to the observed differences in behavior of men and women leaders. Part (a) describes self-selection. Candidacy is a decision made solely by the participant over whether to stand for election. We expect that more competitive women are more likely to *self-select* into candidacy, compared to less competitive women and, further, that the effect of competitiveness on selection into candidacy to be larger for women than for men. Women more starkly sort themselves into candidacy based on competitiveness; for men, competitiveness correlates less strongly with candidacy.

Part (b) of Hypothesis 2 describes what occurs *after* a participant chooses to become a candidate. Both a decision made by the participant, i.e. what campaign message to write, and a decision made by the group, i.e. which candidate to vote for, affect the likelihood of winning. Part (b) argues that more competitive women choose campaign messages that improve their chances of winning. Again, we expect this relationship to be stronger for women than for men.

Empirically, we consider two features of campaign messages: their length and their content. Message length is a type of effort a candidate can exert as she tries to win election. More competitive women who try harder to win election might write longer messages. This makes them more likely to win, since longer messages are more successful in elections.[26] It is also possible that more competitive women are more likely to choose appealing campaign messages, improving their chances of winning election. Since we do not have strong theoretical priors about which submechanism is the most appropriate explanation, we evaluate both. We establish evidence consistent

[26]This is an empirical observation from our data, not a blanket statement about campaigning.

with part (b) of Hypothesis 2 and consider each mechanism in turn.

4 Findings

4.1 Are women more hawkish in our data?

As found in previous studies, men ran for and won election more frequently than women. On average, men ran in the election 43% of the time, compared to 31% for women. A group's elected leader was male in approximately 74% of rounds.[27]

Consistent with Hypothesis 1, female elected leaders in our game exerted more costly effort in intergroup contests than their male counterparts. Table 1 shows the average number of tickets bought by men and women in the ICG and DSG. In the DSG, female elected leaders purchased, on average, approximately 82 more tickets per round than male leaders. This is a large increase in spending on conflict. It is over 8% of their group's total, 1,000 point endowment. In contrast, women bought approximately 29 fewer tickets on average than men in the ICG. This is consistent with findings that women are, on average, less hawkish than their male counterparts.

The differences between the ICG and DSG differ by gender, in ways that immediately suggest greater degrees of non-random selection for women than men. In the DSG, male leaders purchased approximately 61 more tickets compared to the pool of all men who made decisions in the ICG. This is a non-trivial increase of approximately 10%. However, the differences between the ICG and DSG is much larger for women than for men. Female leaders in the DSG purchased approximately 172 more tickets, compared to the pool of all women in the ICG, a 32% increase. The difference between the DSG and ICG is approximately 3 times larger for women than men!

Each of these differences, across gender and part of the game (ICG vs DSG), are statistically significant. Table 2 shows estimates from a regression of ticket purchases on (1) an indicator variable for female participants, (2) an indicator variable for whether the ticket purchase was in the DSG, and (3) the interaction between the two. Since prize values vary by round, and since partici-

[27] Third section of Appendix F shows summary statistics for the distribution of candidates and leaders by gender.

pants select into candidacy before knowing the prize value, we also account for the possibility that differences were driven by which rounds men and women ended up being leaders. In column 2 of Table 2, we add prize-value fixed effects which capture the effect of the prize value itself on ticket levels.[28] The interaction term between DSG and female is statistically significant. While both men and women purchased more tickets as leaders in the DSG compared to the ICG, this difference was much greater for women than for men.

Table 1: Tickets Bought by Men and Women, by Round

Round	Overall (N)	Men (N)	Women (N)
ICG	549:8	561:5	532:6
	(1800)	(1068)	(732)
DSG	644:2	622:8	704:3
	(240)	(177)	(63)

Note: Averages are for all rounds in the ICG and DSG.

[28]For each stage, the 12 prize values are split into 6 prize levels. There are trivial (10 ticket) differences within each prize level, to avoid repeating values exactly in the same stage.

Table 2: Tickets Bought by Men and Women, by Round

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Female	28.923	28.923
	(16.310)	(12.813)
DSG	61.314	57.221
	(27.585)	(21.673)
Female DSG	110.420	126.011
	(52.466)	(41.238)
Constant	561.522	124.143
	(10.401)	(15.542)
Prize FEs?	N	Y
Observations	2,040	2,040
F Statistic	7.391	163.024
Note:	p<0.1; p<0.05; p<0.01	

4.2 Selection Effects

Hypothesis 2 has two components: (a) self-selection into candidacy and (b) the likelihood of winning an election, conditional on having chosen candidacy. Here, we examine how the overall differences between male and female leaders are driven by each channel.

Table 3: Differences in ICG behavior, by Gender

(I) Leaders vs. Not Leaders				
	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Ave. ICG Tickets	565	557	670	520
difference		+8		+150
Ave. ICG NMV	539	507	958	360
difference		+32		+598
(II) Candidates vs. Not Candidates				
	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Ave. ICG Tickets	560	557	616	495
difference		+3		+121
Ave. ICG NMV	520	506	741	259
difference		+14		+482
(III) Winning vs. Losing Candidates				
	Men		Women	
	winner	*loser*	*winner*	*loser*
Ave. ICG Tickets	561	557	675	596
difference		+4		+79
Ave. ICG NMV	522	508	979	662
difference		+14		+317

Note: In the bottom panel (III), winning candidates exclude those who ran unopposed.

As an initial look, the top section of Table 3 compares the two measures of competitiveness for leaders and non-leaders, split by gender. Recall that our two measures are the participant's average ticket purchases in the ICG rounds (Ave. ICG Tickets) and an estimate of the participant's

average estimated non-monetary value to winning based on the ICG rounds (Ave. ICG NMV). For each round in the DSG, we recorded the competitiveness measure of the two groups' leaders and the same measure for all of the non-leaders. The top panel of the table reports the mean of these values across all rounds of the DSG – the average competitiveness of all leaders in the DSG, compared to average competitiveness of non-leaders. This comparison accounts for the fact that some participants became leader more often than others by weighting each participant's competitiveness measure accordingly.[29]

Among men, there is only a small difference in both competitiveness measures between leaders and non-leaders. Male leaders only purchased 8 more tickets per round, on average, in the ICG, compared to male non-leaders. The NMV measure also differs little across leaders and non-leaders for men: the average estimated NMV is only 32 points higher for leaders. The differences in competitiveness are much larger for women. Female leaders purchased 150 more tickets on average in the ICG compared to female non-leaders. Female leaders had an estimated NMV that was 598 points higher than female non-leaders.

Selection into Candidacy

To isolate the self-selection channel, Hypothesis 2 (a), we compare the competitiveness of candidates and non-candidates. The middle portion of Table 3 shows our competitiveness measures, broken down by candidacy and gender.[30] The same patterns found in the top portion of Table 3 are again apparent.

Among men, there are very small differences between candidates and non-candidates. Male candidates bought only 3 more tickets, on average, in the ICG, compared male non-candidates (560 vs. 557). The estimated NMV was only 14 points higher than for male non-candidates. Among women, however, these differences are striking. Female candidates purchased an average of 121

[29]In other words, a participant's competitiveness measure "counts" towards the leader mean in the rounds that she is leader, and towards the non-leader mean in rounds where she is not. To ensure that a few participants who run/win many times do not drive our results, we replicate the analyses excluding the most frequent men and women leaders in Appendix F. Our findings are similar.

[30]Here again, we account for the fact that some participants self-select into candidacy more often than others by weighting each participant's competitiveness measure accordingly.

more tickets in each round of the ICG, compared to female non-candidates (616 vs. 495). Their estimated NMV was over 480 points higher than that of female non-candidates. Self-selection effects are much larger for women than men.

These differences are again statistically significant. In Table 4, we regress an indicator variable for whether the participant was a candidate on their previous ticket purchase average, an indicator for whether they were female, and the interaction between the two. We estimate the regression using 100s of tickets for the ICG average to scale up the coefficients; this decision does not affect the meaning of the results. The unit of observation is the participant-round, for each round of the DSG.

Table 4 shows the results. Looking at Column 1, higher ticket purchases in the ICG were positively associated with the likelihood of candidacy, though this was insignificant. This constituent term describes the effect of ICG purchases on probability of candidacy for men. However, the interaction term shows that - for women - there is a strong, significant effect of ICG ticket purchases on candidacy. Column 2 demonstrates the same patterns using the NMV measure.

The top panel of Figure 1 shows how the predicted probability of being a candidate in the DSG changes for women and men, as their average ICG ticket purchase changes. For men, there is a minimal relationship. For women, the line is steeply upward sloping. Women who purchased approximately 500 tickets per round in the ICG were almost half as likely to stand for candidacy as men who purchased the same amount. In contrast, the most competitive women are predicted to run at higher rates than the most competitive men.

Selection via Election

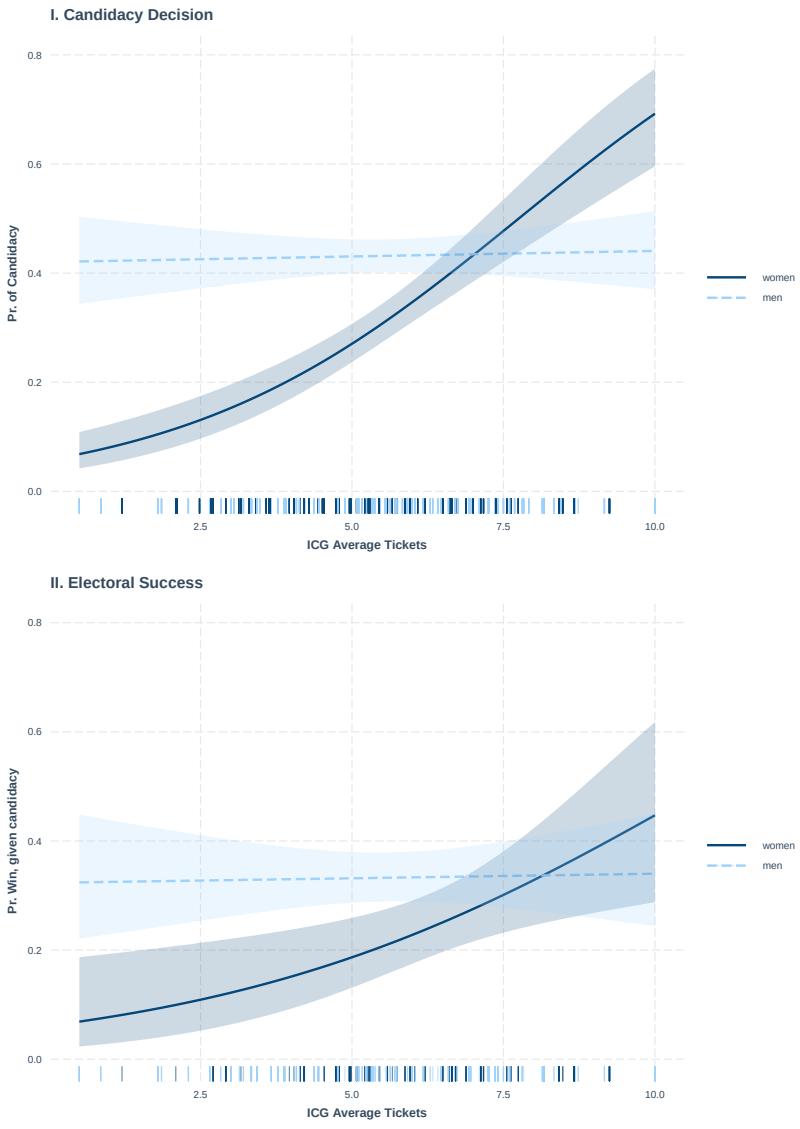
Once a participant self-selects into candidacy, there are also potential selection effects via the election itself, as in Hypothesis 2 (b). The bottom portion Table 3 shows our competitiveness measures, for men and women, broken down by winning candidates versus losing candidates. In other words, these differences show selection effects that result from the election, conditional on the participant having chosen to run. We exclude rounds where a candidate ran unopposed, since

Table 4: Effect of Competitiveness on Candidacy, by Gender

	Dependent variable:	
	Candidate	
	(1)	(2)
Female	2.473	0.958
	(0.336)	(0.130)
Ave. ICG Tickets	0.008	
	(0.031)	
Female Ave. ICG Tickets	0.352	
	(0.056)	
Ave. ICG NMV		0.002
		(0.008)
Female Ave. ICG NMV		0.088
		(0.014)
Constant	0.321	0.286
	(0.183)	(0.074)
Observations	1,786	1,786
Log Likelihood	1,141.795	1,141.795

Note: p<0.1; p<0.05; p<0.01. ICG ticket averages are given in 100s of tickets.

Figure 1: Predicted Effect of Average ICG Ticket Purchases, by Gender



Note: The top panel shows predicted probability of being a candidate in a particular round of the DSG. The results are from Column 1 of Table 4. The bottom panel shows the probability of winning an election in a round of the DSG, conditional on being a candidate. The results are from Column 1 of Table 5. Both plots include ticks on the bottom to show the observed competitiveness of men (light blue) and women (dark blue).

the only selection effect in those rounds was from the candidate's choice to run.

Election-winning men were virtually identical to losing male candidates. Winning male candidates purchased an additional 4 tickets in the ICG, compared to losing candidates. They also had a 14-point higher estimated NMV. The differences for women were much larger, though they are smaller than the differences found when comparing female candidates and non-candidates in the preceding section. Winning female candidates purchased an average of 79 more tickets per round in the ICG and had an estimated NMV that was 317 points higher.

Table 5 shows that these results, too, are statistically significant. This table shows results from regressing whether a candidate won on previous ticket purchases. The unit of observation is the candidate-round, for each round of the DSG. We again exclude observations where a group only had one candidate. As the interaction terms show, the effect of previous ticket purchases on winning is positive for women, and this is stronger than the comparable effect for men.

The bottom panel of Figure 1 shows how the predicted probability of winning an election, conditional on being a candidate, changes for women and men, as their average ICG ticket purchase changes. For men, there is again a minimal relationship between then two. For women, the line is again steeply upward sloping. Women who purchased approximately 500 tickets per round in the ICG were almost half as likely to win election. Only the most competitive women won elections at comparable rates as men.

Campaign Message Content and Length

What explains the gender differences in selection stemming from the election? Here, we show how the decisions made by more competitive female candidates contributes to this difference. Female candidates who were more competitive in the ICG wrote more persuasive campaign messages and invested more effort by writing longer campaign messages. To assess the content of campaign messages, we first coded all messages into 11 categories. Since messages can overlap more than one category, we coded each message for its primary category and also whether any part of the message fell into a particular category. Example categories include:

Table 5: Effect of Competitiveness on Electoral Success, by Gender

	Dependent variable:	
	Victory	
	(1)	(2)
Female	1.991	0.940
	(0.686)	(0.288)
Ave. ICG Tickets	0.008	
	(0.049)	
Female Ave. ICG Tickets	0.244	
	(0.104)	
Ave. ICG NMV		0.002
		(0.012)
Female Ave. ICG NMV		0.061
		(0.026)
Constant	0.739	0.706
	(0.291)	(0.121)
Observations	645	645
Log Likelihood	388.531	388.531

Note: p<0.1; p<0.05; p<0.01. ICG ticket averages are given in 100s of tickets.

- *Track record*, where the candidate appeals to successful past results (Eg "I won the last three rounds where I was leader and can do it again.")
- *Strategy* where the candidate directly states what strategy she would use (Eg "I will bet approximately 100 tickets for each 500 in the prize...")
- *Out-team* The candidate made a negative appeal against the other group (Eg "let's take down that pesky orange team!")

We describe these in particular because they were three most successful types of messages.[31] In our sample, more competitive women in the ICG were more likely to choose one of these three, more successful messages. To show this, we constructed a binary indicator for whether the candidate's primary message fell into one of those three categories. We also constructed a similar indicator for whether any part of the message fell into those categories. We then regressed those indicators on ICG purchases and the NMV measure, an indicator for female participants, and the interaction(s) between the two.[32]

Table 6 shows these results. The outcome variable in columns 1-2 is an indicator for whether the primary message was in one of the three successful categories. The first column uses the ICG ticket measure and the second uses the NMV measure. Columns 3-4 replicate those regressions using the indicator for whether any part of the message fell into those categories as the outcome variable. For men, the relationship between ICG purchases and the likelihood of choosing a successful message type was negative. For women, the relationship was positive and the difference in this effect across genders was significant (per the interaction term). Women who exerted more effort in the ICG were more likely to choose a message whose primary content was one of the three most successful message types. The same is true using the NMV measure, and the same patterns obtain using an outcome variable based on any part of the message.

[31] Appendix C fully describes the coding and categories. It also shows the results for most successful campaign strategies. We did not see evidence that any messages conveyed information about gender.

[32] For analysis of campaign message content and length, we include observations where the candidate ran unopposed, because the candidate wrote their message *before* knowing whether there were other candidates. This distinction does not affect results.

Table 6: Selection of top three campaign strategies, by gender

	Dependent variable:			
	Primary message		Part of message	
	(1)	(2)	(3)	(4)
Female	2.612	1.425	2.420	0.882
	(0.847)	(0.354)	(0.614)	(0.250)
Ave. ICG Tickets	0.115		0.239	
	(0.052)		(0.050)	
Female Ave. ICG Tickets	0.276		0.357	
	(0.128)		(0.097)	
Ave. ICG NMV		0.029		0.060
		(0.013)		(0.012)
Female Ave. ICG NMV		0.069		0.089
		(0.032)		(0.024)
Constant	0.410	0.905	0.770	0.259
	(0.298)	(0.122)	(0.286)	(0.114)
Observations	676	676	676	676
Log Likelihood	336.910	336.910	417.818	417.818

Note: p<0.1; p<0.05; p<0.01. ICG ticket averages are given in 100s of tickets.

In addition to choosing more effective campaign strategies, more competitive female candidates also wrote longer campaign messages. This is consistent with their exerting greater effort to win elections, since writing a longer message takes more time and cognitive effort.[33] Longer messages also contributed to electoral success, as length of campaign message was strongly associated with the likelihood of winning ($p<0.01$).[34]

Table 7 demonstrates that more competitive women wrote longer campaign messages, measured by the number of characters. Participants wrote messages that were an average of 64 characters. Men and women candidates wrote messages of similar length, with averages of 66 and 58 characters, respectively. However, for women, competitiveness was strongly and positively associated with writing longer campaign messages. The most competitive women wrote messages that were about 50% longer than those written by the least competitive women, and were slightly longer than those of the most competitive men.

The correlation between competitiveness in the ICG and (a) self-selection into candidacy and (b) the probability of winning among women *both* contributed to the ultimate difference in observed ticket purchases by men and women leaders in the DSG. Each channel contributed to the fact that women leaders purchased more tickets than men in the DSG. The data suggest that the first channel - self-selection - comprised a larger portion of the overall gender difference, relative to the second channel - winning an election. The differences, for women relative to men, between eventual candidates and non-candidates were larger than those same differences for winning versus losing candidates.

Alternative Explanations and Robustness Checks

We focus on how competitiveness affects candidacy and campaign decisions. However, existing research cites self-confidence as an additional source of gender difference in political engage-

[33] Anson (2018) also uses response length to an open-ended question as a measure of effort from Mturk workers.

[34] Appendix C also shows these results. Longer messages may have signaled higher candidate effort as well as higher quality: Blumenstock (2008) finds that a simple word count outperformed more complex methods in measuring the quality of Wikipedia articles.

Table 7: Length of Campaign Messages, by Effort and Gender

	Dependent variable:	
	Characters in Campaign Message	
	(1)	(2)
Female	38.598	16.420
	(10.531)	(4.261)
Ave. ICG Tickets	2.345	
	(0.876)	
Female Ave. ICG Tickets	5.154	
	(1.673)	
Ave. ICG NMV		0.586
		(0.219)
Female Ave. ICG NMV		1.288
		(0.418)
Constant	79.443	69.352
	(5.248)	(2.159)
Observations	676	676
F Statistic (df = 3; 672)	5.919	5.919

Note: p<0.1; p<0.05; p<0.01. ICG ticket averages are given in 100s of tickets.

ment.[35] Such differences cannot explain our results.[36] In the appendix we measure confidence by using participants' average payoffs in the Individual Contest Game.[37] The participants' payoffs give them concrete feedback about how their strategies succeed or fail in scoring points and how this compares to the payoffs of their opponents. We analyze whether this measure of confidence explains decisions to run, the likelihood of winning, and subsequent ticket purchases. We do not find evidence of any gendered differences in these relationships. A participant's average ICG payoff is positively correlated with candidacy and electoral success, but its effect is very similar across genders. We can also show that gender differences in confidence in ability to win elections do not explain our results. If this were the case, we would expect a winning (losing) electoral history would increase (decrease) a participants' confidence in their ability to win elections, and that this effect would vary across gender. We do not find a significant gender gap in the effect of past electoral performance on the decision to become a candidate or the likelihood of becoming a leader. This is consistent with recent observational work showing that election losses have similar dissuading effects across gender.[38]

The appendix addresses several other alternative explanations. We examine the possibility that a desire to win re-election, which occurs in odd-numbered rounds before groups are reshuffled, affected results and that re-election effects differ by gender. We find evidence of neither. We then assess whether the election effect - wherein the experience of being elected induces people to buy more tickets - differed by gender. We find similar election effects for men and women. We also repeat all these analyses using data from rounds when participants were randomly selected as group leader to measure competitiveness. The results are similar.[39] Finally, we redo analyses excluding rounds in which the prize value was lower than participants' ticket endowments. We find no significant differences between these results and the above analysis, which includes low

[35] Fox and Lawless (2011); Wolak (2020).

[36] Note that Kanthak and Woon (2015) design their experimental protocol to directly isolate confidence versus election aversion. They too conclude that competitiveness is the stronger explanation.

[37] Appendix G.

[38] Bernhard and de Benedictis-Kessner (2020).

[39] Appendix F.

prize values rounds.[40]

5 Conclusion

Female chief executives may choose more hawkish policies - at least in part - because of their own preferences, world views, or tolerance for conflict and competition. To become a chief executive, one must choose to run and then win a competitive election. Neither aspect of this gauntlet is for the faint of heart. As a result, the very things that determine whether a woman chooses to run mean that those women who, in Margaret Thatcher's words, "get through" are far less fazed by conflict and competition than the average woman. While the choices of Iron Ladies are undoubtedly shaped by their strategic environments, to quote another famous Lady (Gaga), many were simply "born this way."

Consistent with this, we demonstrate that female elected leaders invest more in costly competition than their male counterparts in a controlled laboratory experiment. Women who choose to run and then eventually win elections are among the most competitive, and the difference between them and those who do not choose to run or are unsuccessful in elections is much larger than analogous differences for men. We found these gendered patterns despite making participants anonymous, which ameliorates the effects of overt gender-bias and stereotyping.

Our research highlights two channels for future research. First, our approach does not allow us to compare the relative strength of selection mechanisms versus other real-world mechanisms based on stereotyping. We think both are likely present, but to isolate the former, we chose a protocol that downplayed the latter. A natural extension would be to repeat our experiments, but remove anonymity and allow participants to know (or at least presume) the gender identity of their fellow participants. We suspect that removing anonymity would magnify our findings. Since only the most competitive women run in and win election in our anonymous game, we hypothesize that removing anonymity would exacerbate these selection effects.

[40] Appendix E.

Second, we chose a conflictual context that resembles many intergroup interactions and decisions made by leaders. Many intergroup interactions are cooperative and non-competitive. Studies of intragroup public goods games find that women, on average, donate twice as much to their anonymous partners in a dictator experiment.[41] Endogenous leader selection also helps select more pro-social individuals who choose policies closer to the group's optimum.[42] Initial evidence suggests that women are equally willing to lead, compared to men, in settings they believe to be cooperative, rather than conflictual.[43] Self-selection and electoral effects by gender may be weaker in settings that are less competitive, and where incentives across groups are more aligned.

[41] Eckel and Grossman (1998).
[42] Hamman, Weber and Woon (2011).
[43] Hong et al. (2014).

References

Andreoni, James and Lise Vesterlund. 2001. "Which is the fair sex? Gender differences in altruism." *The Quarterly Journal of Economics* 116(1):293–312.

Anson, Ian G. 2018. "Taking the time? Explaining effortful participation among low-cost online survey participants." *Research & Politics* 5(3):2053168018785483.

Anzia, Sarah and Christopher Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55(3):478–493.

Barnes, Tiffany D and Diana Z O'Brien. 2018. "Defending the realm: The appointment of female defense ministers worldwide." *American Journal of Political Science* 62(2):355–368.

Barnhart, Joslyn N., Robert F. Trager, Elizabeth N. Saunders and Allan Dafoe. 2020. "The Suffragist Peace." *International Organization* 74(4):633–670.

Bendix, William and Gyung-Ho Jeong. 2020. "Gender and Foreign Policy: Are Female Members of Congress More Dovish than Their Male Colleagues?" *Political Research Quarterly* 73(1):126–140.

Bernhard, Rachel and Justin de Benedictis-Kessner. 2020. "The Boys—and Girls—Are Back in Town: Male and Female Candidates Are Similarly Dissuaded by Losing Elections.".

Bernhard, Rachel, Shauna Shames and Dawn Langan Teele. 2020. "To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office." *American Political Science Review* pp. 1–16.

Blumenstock, Joshua E. 2008. Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web*. pp. 1095–1096.

Campbell, Anne. 2013. *A mind of her own: The evolutionary psychology of women*. OUP Oxford.

Caprioli, Mary. 2000. "Gendered conflict." *Journal of peace research* 37(1):51–68.

Caprioli, Mary and Mark A Boyer. 2001. "Gender, violence, and international crisis." *Journal of Conflict Resolution* 45(4):503–518.

Chaudoin, Stephen and Jonathan Woon. 2018. "How hard to fight? Cross-player effects and strategic sophistication in an asymmetric contest experiment." *The Journal of Politics* 80(2):585–600.

Clayton, Amanda and Pär Zetterberg. 2018. "Quota shocks: Electoral gender quotas and government spending priorities worldwide." *The Journal of Politics* 80(3):916–932.

Croco, Sarah E and Scott Sigmund Gartner. 2014. "Flip-flops and high heels: An experimental analysis of elite position change and gender on wartime public support." *International Interactions* 40(1):1–24.

Dube, Oeindrila and SP Harish. 2020. "Queens." *Journal of Political Economy* 128(7):2579–2652.

Eckel, Catherine C and Philip J Grossman. 1998. "Are women less selfish than men?: Evidence from dictator experiments." *The economic journal* 108(448):726–735.

Eichenberg, Richard C. 2016. "Gender difference in American public opinion on the use of military force, 1982–2013." *International Studies Quarterly* 60(1):138–148.

Enloe, Cynthia. 1989. *Bananas, beaches and bases: Making feminist sense of international politics*. Univ of California Press.

Folke, Olle and Johanna Rickne. 2016. "The glass ceiling in politics: Formalization and empirical tests." *Comparative Political Studies* 49(5):567–599.

Folke, Olle, Johanna Rickne and Daniel M Smith. 2021. "Gender and dynastic political selection." *Comparative Political Studies* 54(2):339–371.

Fox, Richard and Jennifer 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.

Fox, Richard L and Jennifer L Lawless. 2010. "If only they'd ask: Gender, recruitment, and political ambition." *The Journal of Politics* 72(2):310–326.

Fox, Richard L and Jennifer L Lawless. 2014. "Uncovering the Origins of the Gender Gap in Political Ambition." *The American political science review* 108(3):499–519.

Goldgeier, James M. 1994. *Leadership Style and Soviet Foreign Policy: Stalin, Khrushchev, Brezhnev, Gorbachev*. Johns Hopkins University Press.

Hamman, John R, Roberto A Weber and Jonathan Woon. 2011. "An experimental investigation of electoral delegation and the provision of public goods." *American Journal of Political Science* 55(4):738–752.

Holman, Mirya R, Jennifer L Merolla and Elizabeth J Zechmeister. 2016. "Terrorist threat, male stereotypes, and candidate evaluations." *Political Research Quarterly* 69(1):134–147.

Holman, Mirya R, Jennifer L Merolla, Elizabeth J Zechmeister and Ding Wang. 2019. "Terrorism, gender, and the 2016 US presidential election." *Electoral Studies* 61:102033.

Hong, Alain, Juliette Schaafsma, Per van der Wijst and Aske Plaat. 2014. Taking the lead: Gender, social context and preference to lead. In *European Conference on Management, Leadership & Governance*. Academic Conferences International Limited p. 445.

Horowitz, Michael C, Allan C Stam and Cali M Ellis. 2015. *Why leaders fight*. Cambridge University Press.

Judge, Dayna. 2021. "Backwards and in Heels: Post-Crisis Female Leadership Signals Competence and Change." *Working paper* .

Kanthak, Kristin and Jon Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." *American Journal of Political Science* 59(3):595–612.

Koch, Michael T and Sarah A Fulton. 2011. "In the defense of women: Gender, office holding, and national security policy in established democracies." *The Journal of politics* 73(1):1–16.

Lawless, Jennifer L. 2004. "Women, war, and winning elections: Gender stereotyping in the post-September 11th era." *Political Research Quarterly* 57(3):479–490.

Niederle, Muriel and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The quarterly journal of economics* 122(3):1067–1101.

Niederle, Muriel and Lise Vesterlund. 2011. "Gender and competition." *Annu. Rev. Econ.* 3(1):601–630.

Post, Abigail S and Paromita Sen. 2020. "Why can'ta woman be more like a man? Female leaders in crisis bargaining." *International Interactions* 46(1):1–27.

Powell, Jonathan and Karina Mukazhanova-Powell. 2019. "Demonstrating Credentials? Female Executives, Women's Status, and the Use of Force." *Journal of Women, Politics & Policy* 40(2):241–262.

Preece, Jessica and Olga Stoddard. 2015. "Why Women Don't Run: Experimental Evidence on Gender Differences in Political Competition Aversion." *Journal of Economic Behavior and Organization* 117(296-308).

Regan, Patrick M. and Aida Paskeviciute. 2003. "Women's access to politics and peaceful states." *Journal of Peace Research* 40(3):287–302.

Reiter, Dan. 2015. "The positivist study of gender and international relations." *Journal of Conflict Resolution* 59(7):1301–1326.

Ryan, Michelle K, S Alexander Haslam and Clara Kulich. 2010. "Politics and the glass cliff: Evidence that women are preferentially selected to contest hard-to-win seats." *Psychology of Women Quarterly* 34(1):56–64.

Schramm, Madison and Alexandra Stark. 2020. "Peacemakers or Iron Ladies? A Cross-National Study of Gender and International Conflict." *Security Studies* pp. 1–34.

Schroeder, Theresa. 2017. "When security dominates the agenda: The influence of ongoing security threats on female representation." *Journal of Conflict Resolution* 61(3):564–589.

Schwartz, Joshua A. and Christopher W. Blair. 2020. "Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining." *International Organization* 74(4):872–895.

Schwarz, Susanne and Alexander Coppock. 2020. "What Have We Learned About Gender from Candidate Choice Experiments? A Meta-Analysis of 67 Factorial Survey Experiments." *Unpublished Manuscript* June 30:1–40.

Sjoberg, Laura, Kelly Kadera and Cameron G Thies. 2018. "Reevaluating gender and IR scholarship: Moving beyond Reiter's dichotomies toward effective synergies." *Journal of Conflict Resolution* 62(4):848–870.

Spisak, Brian R, Peter H Dekker, Max Krüger and Mark Van Vugt. 2012. "Warriors and peacekeepers: Testing a biosocial implicit leadership hypothesis of intergroup relations using masculine and feminine faces." *PloS one* 7(1):e30399.

Swers, Michele. 2007. "Building a Reputation on National Security: The Impact of Stereotypes Related To Gender and Military Experience." *Legislative Studies Quarterly* 32(4):559–595.

Thomas, Melanee and Marc André Bodet. 2013. "Sacrificial lambs, women candidates, and district competitiveness in Canada." *Electoral Studies* 32(1):153–166.

Van Vugt, Mark and Brian R Spisak. 2008. "Sex differences in the emergence of leadership during competitions within and between groups." *Psychological Science* 19(9):854–858.

Wolak, Jennifer. 2020. "Self-Confidence and Gender Gaps in Political Interest, Attention, and Efficacy." *The Journal of Politics* 82(4):1490–1501.

Appendix Sections

Appendix A: Protocol, Recruitment, Demographics

Appendix B: Deception Justification

Appendix C: Campaign Messages Coding, Message Effectiveness

Appendix D: Analysis Based on Random Selection

Appendix E: Dropping Low Value Rounds

Appendix F: Alternative Explanations: Election, Re-Election Effects

Appendix G: Confidence Analysis

A Appendix: Full Protocol Description, Recruitment, Demographics

This section of the appendix first goes through the full experimental protocol. It then describes recruitment via MTurk and the demographics of the participants.[44]

A.1 Protocol

Before playing, participants watched an animated video explaining the rules of the contest game. We hired a graphics designer to ensure that the instructional video was clear and engaging.[45] Participants then answered four quiz questions to ensure that they understood the mapping between their choices and payoffs. Figure A.1 shows screen shots from the instructional video (top two panels) and from the instructions quiz questions (bottom two panels). If players answered incorrectly, they were shown the correct answer and given an explanation. If players answered correctly, they were told that their answer was correct and given the same explanation.

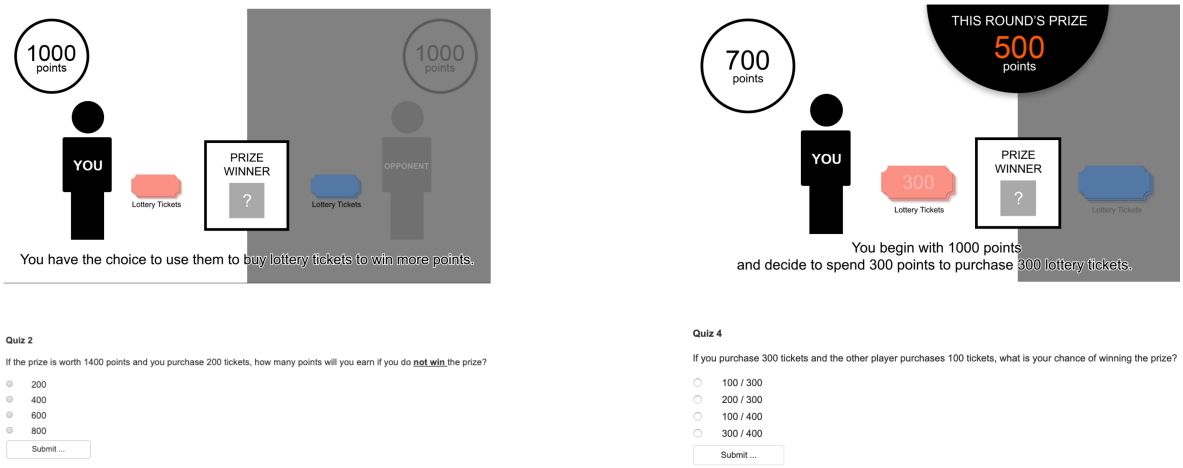


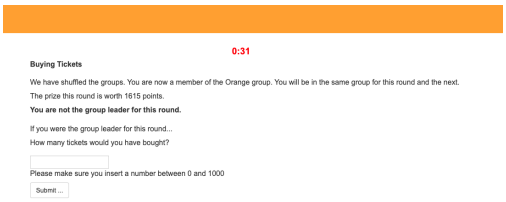
Figure A.1: Screen Captures from Instructional Video and Quiz

The bottom panels of Figure A.2 show examples of a participant's choice to run and their campaign message.

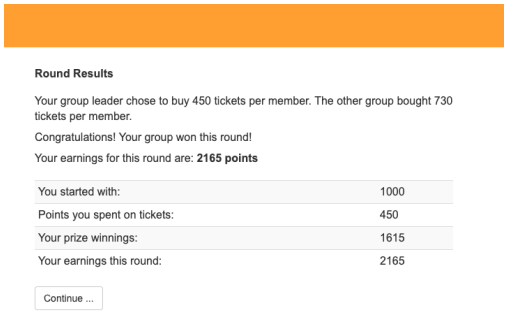
[44]For Appendix A, Appendix B, and Appendix C, we have used the same language as in our other working paper as the underlying experiment is the same.

[45]The video is available here: `https://www.youtube.com/watch?v=3ywZvA0CLy8`.

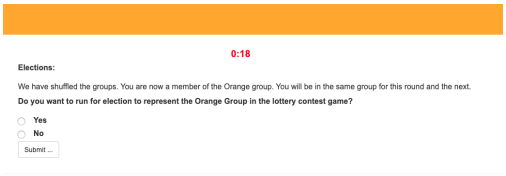
The top two panels of Figure A.2 show screen shots from the game. The top left panel shows an example of a player who was not chosen as group leader and was then asked how many tickets she would have purchased. The screen for a group leader looks similar. The top right panel shows an example of what a player sees at the end of a round. They learned how many tickets that they/their leader purchased, how many tickets the opposing leader purchased, whether they won, and their earnings for that round.[46]



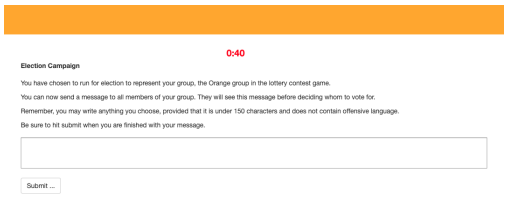
(a) Ticket Purchase



(b) Round Results



(c) Candidate Decision



(d) Campaign Message

Figure A.2: Screen Captures from the Random Selection and Democratic Selection Parts of the Game

Table A.1 shows the prize values for each round and each part of the game. We chose these values to make sure that the prize value was sufficiently enticing to get people to bid, based on test runs. We picked values that weren't exactly round numbers, like 2000, in order to decrease the power of focal points, like bidding exactly one half of the prize value.

We varied the prizes and slightly varied the order across sections to minimize order effects. We created a list of paired prize values that differed from each other by 10 points (eg 2715 and 2705, 1235 and 1225). We then made sure that one value from each pair appeared in the first and second

[46]The red times indicate a countdown timer for each decision, generally one minute. This let us drop players who timed-out or dropped out for other reasons and still keep the game moving. In practice, this rarely happened and respondents did not seem pressed for time.

halves of each section (eg, in the ICG, 2715 appears in round 1 - first half - and 2705 appears in round 9 - second half). This makes it very unlikely that participants would identify a pattern and base their play on anticipated future prize values.

Table A.1: Prize Value by Round and Section

Round	Indiv. Contest Game (ICG)	Random Sel. Game (RSG)	Democratic Sel. Game (DSG)
1	2715	1615	2715
2	275	280	275
3	1235	2475	1235
4	2475	2715	2475
5	2035	2035	2035
6	1605	1235	1605
7	1225	2025	1225
8	2025	2705	2025
9	2705	1605	2705
10	2465	2465	2465
11	280	275	280
12	1615	1225	1615

A.2 Recruitment

We recruited 162 participants for 10 sessions of our game from Amazon's Mechanical Turk (MTurk) in December 2019. Using such online platform has become popular for survey experiments in political science because the online samples tend to be more representative than in-person convenience samples. Berinsky, Huber and Lenz (2012) and Mullinix et al. (2015) have have drawn similar inferences by conducting identical studies over the Mturk and population-based samples. Due to the difficulty of conducting real-time interactive games online, Mturk has been less often used for group games like ours. We overcome this added difficulty of coordinating simultaneous intergroup games using the Software Platform for Human Interaction Experiments, or "SoPHIE" (Hendriks, 2012). This platform allows us to place participants into virtual waiting rooms where they wait for other participants to finished their timed tasks before they are placed into pairs or groups.

Our recruitment procedure ensured that the participants were attentive. Individuals who wished to partake in the study had to fill out a brief pre-survey that was posted no longer than an hour before the start time. They had to pass a reCAPTCHA screen and explicitly agree to show up a pre-designated time before submitting the pre-survey. The individualized study link and instructions were then separately message to those who had completed the survey. They then had to open the message and follow the instructions to enter the session at the agreed-on time. If they didn't show up on time as they had agreed, they would not be able to participate in the game.[47] For those who did show up on time and proceeded to participate in the study, they were instructed and incentivized to focus on the game as each lottery contest game decisions were timed. Participants who didn't make the purchasing decision within one minute were excused from the session and paid a prorated bonus for the time spent on the study. Participants are placed in the virtual waiting room (for upto about a minute) after each round while waiting for others to finish making their purchasing decisions. As they don't know when they will be released from the waiting room and set off to another one-minute task, they had to be attentive in order to successfully complete the

[47] For those who showed up late, we still paid the $5 show up fee even though we dismissed them from the study.

game without being excused.

A.3 Demographics of MTurk Sample

Our sample was closer to the United States national averages on most demographic characteristics compared to most university student samples. Figure A.3 shows demographics in our sample compared to the 2018 Cooperative Congressional Election Study (CCES) sample and an in-person university laboratory sample used by Anderson et al. (2013). The CCES uses matching and post-stratification weighting to construct a nationally-representative sample of American adults. We chose Anderson et al. (2013) as a benchmark for student samples because it collected and reported a larger number of demographic characteristics than most studies using student samples. Compared to the CCES, our MTurk participant pool was on average about 10.5 years younger while the university laboratory pool was about 27.2 years younger. Because of the nature of the setting of typical university lab experiments, the MTurk sample was much closer to the CCES benchmark in the distribution of education levels. Similarly, our MTurk sample has a more representative distribution of income levels than the in-person lab sample.[48] Our MTurk sample had 10.8 percent fewer women than the CCES sample while the in-person lab sample had 9.5 percent more. The racial composition of the university lab sample was closer to CCES benchmark. The student sample had about 7.0 percent more whites than CCES while our MTurk sample had 12.5 percent more.

[48]CCES and our MTurk sample measure household income, and Anderson et al. (2013) ask for the participants' parents' income.

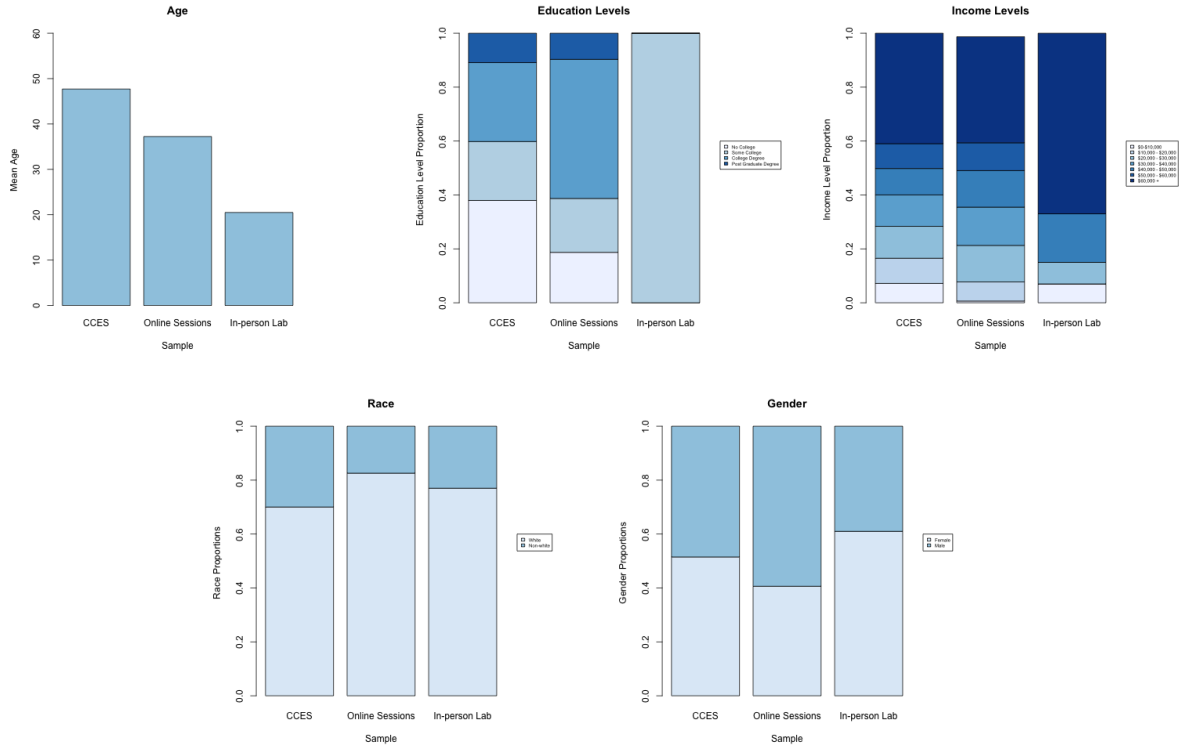


Figure A.3: Comparison of Sample Characteristics

B Appendix: Deception Justification

The experimental protocol used two minimal instances of deception. First, during the Random Selection Game, we chose three rounds to tell every participant that they had been randomly selected to be leader of their group. Second, we told respondents that they were in groups of 7, but we actually used groups of 8 in case there were mid-experiment dropouts and we did not notify respondents if participants had dropped out, which could change their group size.

Here, we first describe why neither instance of deception caused harm to the respondents. Then we describe why alternate protocols would not be feasible and why we chose our particular protocol.

Possible Harm of the Random Leader Deception

There are three reasons why we believe that the potential harms of this decision are negligible.

Pre-brief and Debrief

Our respondents were made aware of the possibility of deception before consenting to participate in the experiment. Pre-brief helps respondents know about the possibility of deception, which allows them to make a partial judgement and opt out of participation. Debrief helps respondents know exactly what we did so that they do not leave wondering about the nature of deception. The informed consent notification stated:

> *As part of this research design, you may not be told everything or may be misled about the purpose or procedures of the research. You will be fully informed about the procedures and any misinformation at the conclusion of the study.*

Additionally, the debrief statement made both instances of deception explicit (emphasis added in bold):

> *Thank you for your participation in this experiment. This study is concerned with the effect of democratic selection on the way leaders play contest games. Major decisions*

about intergroup interactions are made by leaders, and groups vary significantly in how those leaders come to power. Our research assesses how leader selection mechanisms affect intergroup contests, which are a natural analogue to conflictual interactions like war. There is a lack of literature on the effect that different types of leader selection have in the way the selected leaders behave on behalf of their groups. We suspect that democratically elected leaders will exert excessive efforts in intergroup contests at the expense of overall group payoff.

The nature of the study we are conducting required minor deception on our part. Because of the online nature of the experiment, we had to think of ways to handle participant dropouts without disrupting the experiment for other participants. Our main deception was employed in order to address this dropout problem. We began each session with 16 participants. Whenever we formed groups, players were evenly split into two groups. In Part 2 and Part 3, we told you that you have been placed into a group of 7 players, but that might not have been accurate due to participant dropouts. Thus, while we expected groups to have an average of 7 players, it was possible that they had 8, 6, or even 5 members at times. It was also possible that the groups were slightly different sizes if we could not make an even division. However, in all cases, we calculated lottery ticket totals and payoffs as if groups were a standard 7 players.

Moreover, if the selected leader of a group in Part 2 and Part 3 dropped out after being selected but before making a purchasing decision, we used his predicted decision that we estimated using the data of his game playing behavior that we obtained in Part 1.

Lastly, in three of the ten rounds in Part 2, we told all participants that they have been selected as the leader. While leader selection was random in all other rounds of Part 2, it was actually not in those three rounds. We did this in order to make sure

we had data on all participants playing the game as the leaders of their groups. This feature did not have any impact on your earnings because we did not select one of these rounds as the one that bonuses are calculated from.

If you have concerns about your rights as a participant of this study, please contact [contact information].

We did not have any respondents contact us or our institution's IRB to object to this use of deception. Additionally, several websites allow MTurk workers to discuss and rate "Requesters" (people like us who post a task). We have read all of these carefully and have not seen any negative mentions of deception.

Economic or Material Costs

If any respondent felt uncomfortable with the possibility of deception at the informed consent stage, they could decline to participate in the experiment and *still* receive a $5.00 show up fee. At the point of declining to participate, respondents would have interacted with our MTurk task for less than two minutes. No respondents declined to participate.

Even after agreeing to participate, the economic or material costs were zero. In calculating payment, we did not use the three rounds in which we had told all participants that they had been randomly selected. Therefore, even if they had changed their choices based on the deception, it would not affect their compensation.

Cognitive or Psychological Trauma

We do not think there is any risk of a respondent feeling traumatized either by the experience of the game or upon learning that they had been deceived. In some experiments, deception entails doing something that could cause the respondent to doubt factual information in the broader world or question their self-worth. We cannot think of any reason why a respondent would feel trauma upon learning that she was not the leader of a group during three rounds of a lengthy game or that a group member had dropped out and her group did not consist of exactly seven members.

Alternative Protocols and Why We Didn't Use Them

We carefully considered alternative protocols and ultimately decided on the protocol described here because the costs of an alternative protocol outweighed the negligible harms.

Elicit Strategies Protocol

Some experiments ask respondents to pre-commit to a strategy *before* they find out whether they are the leader of a group or before some other "shock" is realized. This would not have worked for our experiment for theoretical reasons.

Our theory focuses on the effect *of becoming group leader*, either through random selection or democratic election. If we think of the experience of becoming group leader as the "treatment" then eliciting strategies before treatment would not work. In practice, we found differences between answers to the hypothetical question "what would you have done if you were leader" and participants' actual choices, which is suggestive that elicited strategies would not be the same as the observed strategies, chosen after the "treatment" of becoming leader.

If we asked respondents to choose strategies before knowing whether they were the leader of a group (either through random selection or through the election protocol), then we would not be able to assess the effect of becoming the leader of your group on your effort level in the contest. Choosing strategies before the respondent is actually chosen as group leader would mean that the outcome of interest - effort exerted in a contest - is temporally prior to the treatment of interest, which occurs when a respondent is chosen as leader of her group.

Full Randomization

Another alternative protocol would have simply let a randomization device select all group leaders in the Random Selection Game, without us making everyone a leader in certain rounds. Here, the major downside is that we would not get data on leadership decisions for all respondents.

Figure B.1 shows the binomial distribution when the probability of success (being chosen leader) is 1/8, with 12 draws. For a particular respondent, there is a 20% chance (approximately)

that they will never be leader, 35% chance of being leader once, 27% chance of being leader twice, and 18% chance of being chosen 3 or more times.

The problem compounds when considering that there are only two low-value rounds, which we used in our analysis of the non-monetary value to winning. For any given participant, their chances of being chosen leader in at least one of those two rounds is only 23%, meaning that we would lack low-value round data for over 75% of the participants.

Figure B.1: Binomial Distribution for Random Leader Selection

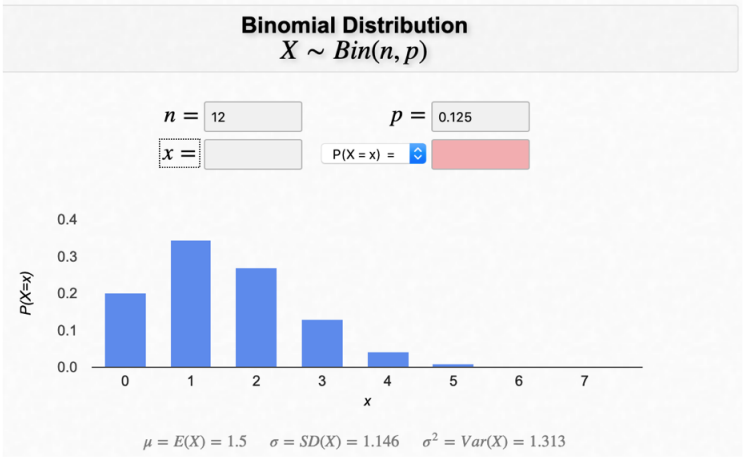


Figure Source: Bognar, Matt. Department of Statistics and Actuarial Science, University of Iowa. `https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html`.

The subject payments for these experiments were approximately $3,300 in total, for a final sample that included 162 participants. If we used a truly random leader assignment, we would need to recruit 704 participants, in expectation, to have data from at least one low value round for each participant. That would cost approximately $14,340 which is a huge amount for this type of research. This is also an under-estimate of the total cost to collecting data equivalent to ours, because this would be enough to get one data point for each respondent, whereas we currently have more than one. Additionally, although this number of participants would get one data point in the RSG for each participant, in expectation, there is no guarantee that those with one or more RSG data points would be elected leaders in the DSG, which would leave us unable to make within-

participant comparisons.

C Appendix: Campaign Messages Coding

This section of the appendix first describes how we coded the content of campaign messages. It also gives data about the prevalence and correlation among message types. We then show results that describe which campaign messages were most effective. This information was used in the main manuscript to describe whether certain participants were more likely to choose the more successful types of campaign messages, in terms of the message content.

C.1 Coding content of messages

The DSG entails campaign messaging from candidates to voters. Here, we code every campaign message into 11 categories and describe some important patterns. A full codebook is available on request, but here, we give a brief summary and example of each type of message.

- Appeals to candidate skill (*skill*)
  - The candidate made an appeal to his or her skill in the game.
  - Eg "I understand the game well and will get us the most money."
- Strategy promises (*strategy*)
  - The candidate made a statement about what strategy she would use.
  - Eg "I will bet approximately 100 tickets for each 500 in the prize..."
- Appeals to track record (*trackrecord*)
  - The candidate made a positive appeal to her past outcomes.
  - Eg "I won the last three rounds where I was leader and can do it again."
- Appeals against previous leaders (*pastbad*)
  - The candidate made a negative appeal against past leaders' decisions.

  - Eg “Dont pick that dude again. I can win for us.”

- Appeals to in-team identity (*inteam*)
  - The candidate made an appeal to in-group identity, usually by referencing “we,” “us,” or “our group.”
  - Eg “Hi all! I promise to make the best choices for our group if you elect me.”

- Appeals against the out-team (*outteam*)
  - The candidate made a negative appeal against the other group.
  - Eg “let’s take down that pesky orange team!”

- Appeals to higher bidding (*bidhigh*)
  - The candidate made an appeal for higher or riskier bidding.
  - Eg “I like higher bids go big or go home”

- Appeals to lower bidding (*bidlow*)
  - The candidate made an appeal for lower or more conservative bidding.
  - Eg “i wont go over 200 we will get 800 at the least.”

- Appeals to speed (*speed*)
  - The candidate promised to make quick decisions. Mturkers value their time.
  - Eg “I will make fast decisions to get this game moving faster.”

- Humor (*humor*)
  - The candidate used humor.
  - Eg “baby yoda for president”

- Null (*null*)

- These messages didn't really have much content or seemed to be mistakes/typos.
- Eg "Vote for me"

We coded the primary category of each message, and we also constructed a set of indicator variables for whether the message contained any of a particular type of content. For example, a message might primarily consist of an appeal to the candidate's successful track record (primary = *trackrecord*), but it might also contain humor (*msg trackrecord* = 1 and *msg humor* = 1).

Table C.1 shows the prevalence of each type of message. The first two columns focus on the primary message type only. Column 1 shows the frequency of each primary message type and the percentage of all messages that were primarily of that type. Column 2 repeats those summary statistics, but only limited to the messages that won that particular election. Skill messages were the most frequently occurring, followed by those using humor or appealing to the candidate's track record.

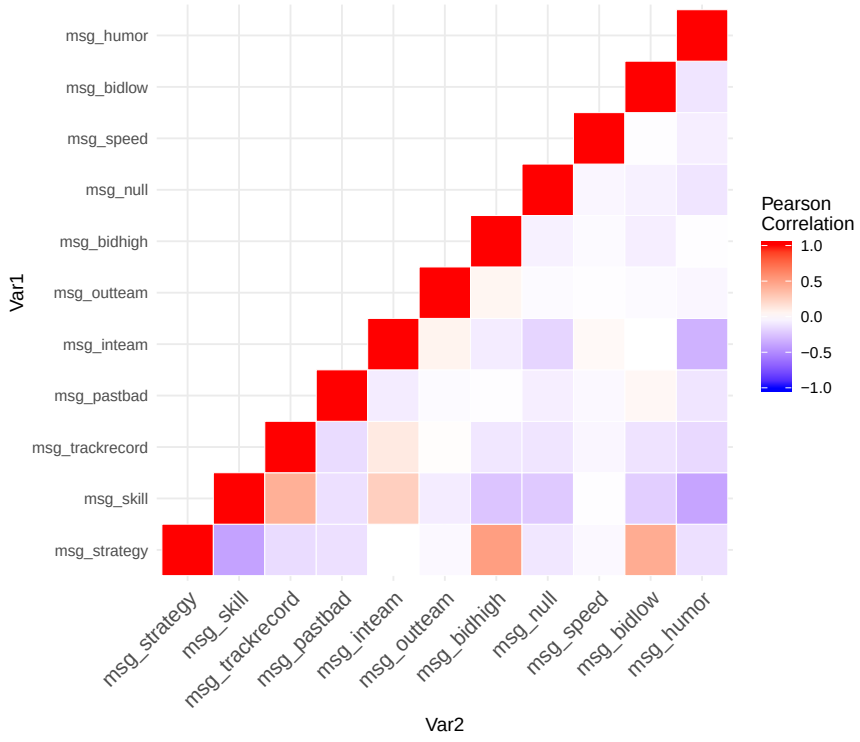
Columns 3 and 4 repeat those summary statistics, but they include all of a message's content, not just its primary message type. The relative prevalence of most categories is similar to columns 1 and 2 with two main exceptions. The prevalence of messages with any *inteam* content increases. This is because this category was broad, capturing the use of keywords like "Let's" and "us." The prevalence of *strategy* also increased. This is because other categories are nested in this one. For example, any appeal to bidding lower (*bidlow*) also necessarily implies an appeal to a particular strategy.

Figure C.1 shows the correlation coefficients for each type of message content, for those messages that contained more than one type of content. There aren't any particularly striking patterns, with most pairwise correlation coefficients fairly low. Strategy messages were positively correlated with bidhigh and bidlow, by construction as explained above. Skill messages were more likely to include appeals to a candidate's track record and inteam verbiage, but less likely to contain humor.

Table C.1: Election success rates by message type

	Primary Message Only		Any Message Content	
	Prevalence (N)	Win. Msg. Only (N)	Prevalence (N)	Win. Msg Only (N)
skill	0.26 (174)	0.27 (62)	0.51 (343)	0.60 (140)
humor	0.16 (109)	0.09 (20)	0.20 (134)	0.13 (31)
trackrecord	0.15 (104)	0.25 (58)	0.18 (120)	0.29 (67)
inteam	0.11 (72)	0.08 (18)	0.38 (260)	0.45 (104)
pastbad	0.08 (56)	0.08 (19)	0.09 (64)	0.09 (22)
bidhigh	0.06 (41)	0.07 (16)	0.07 (49)	0.08 (18)
strategy	0.05 (36)	0.10 (23)	0.17 (113)	0.20 (47)
null	0.05 (34)	0.01 (3)	0.05 (34)	0.01 (3)
bidlow	0.05 (33)	0.03 (6)	0.06 (41)	0.03 (8)
speed	0.02 (13)	0.02 (5)	0.03 (21)	0.03 (7)
outteam	0.01 (4)	0.01 (3)	0.01 (4)	0.01 (3)

Figure C.1: Correlation of message content



Note: Figure shows the correlation coefficients for each type of message content, for messages that had content pertaining to more than one type of content.

C.2 Campaign Message Effectiveness

What campaign messages - in terms of message content and length - were most successful?

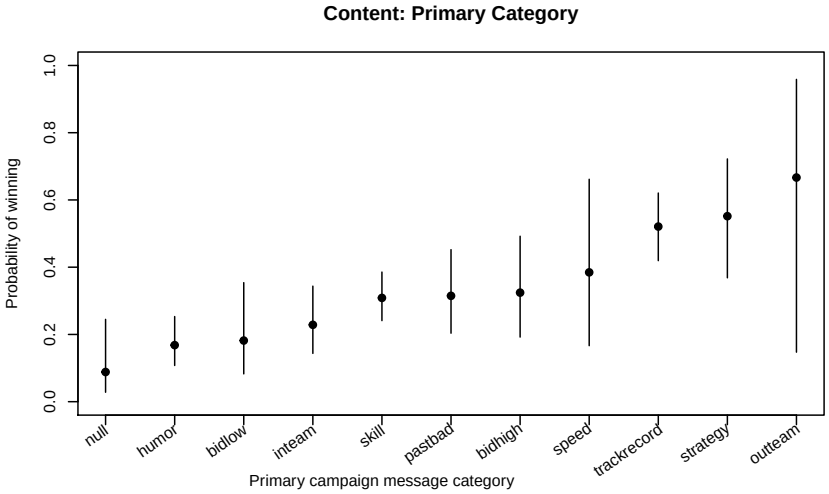
Table C.2 estimates the likelihood of winning election based on message content and length. Columns 1 and 2 use indicator variables for each type of message content as the regressors. The base category that is withheld is the content category labelled "null" in the previous section, corresponding to messages that didn't really have discernible content. Column 1 codes these indicators based on the primary message content. Column 2 codes these indicators based on whether any part of the message fell into a particular category. In both cases, the three most successful categories are "trackrecord," "strategy," and "outteam." Column 3 uses message length, in terms of characters, as the explanatory variable. Longer campaign messages are associated with a higher probabilty of victory.

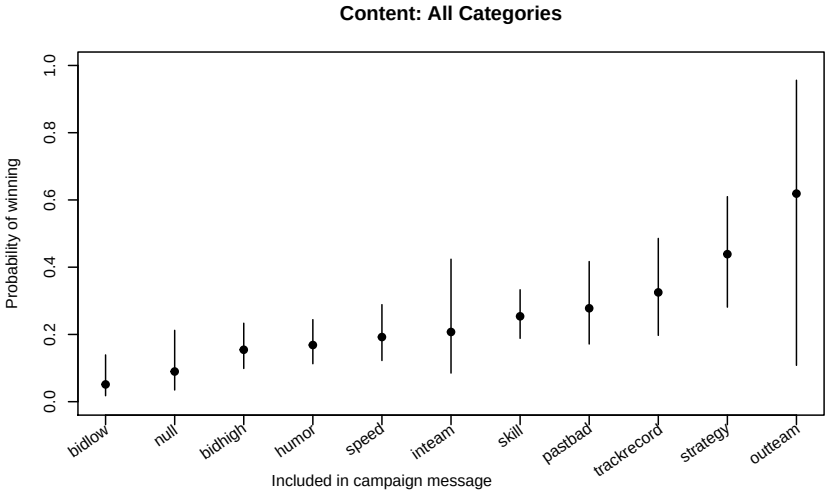
Figure C.2 shows the estimates visually. The top panel shows the predicted likelihood of winning the election for each primary message type. The message types are ordered from least successful (left hand side) to most successful (right hand side). The middle panel repeats this, but using the estimates from Column 2, based on whether a message's content fell into any of the particular categories. The bottom panel shows the relationship between message length and the probability of winning an election.

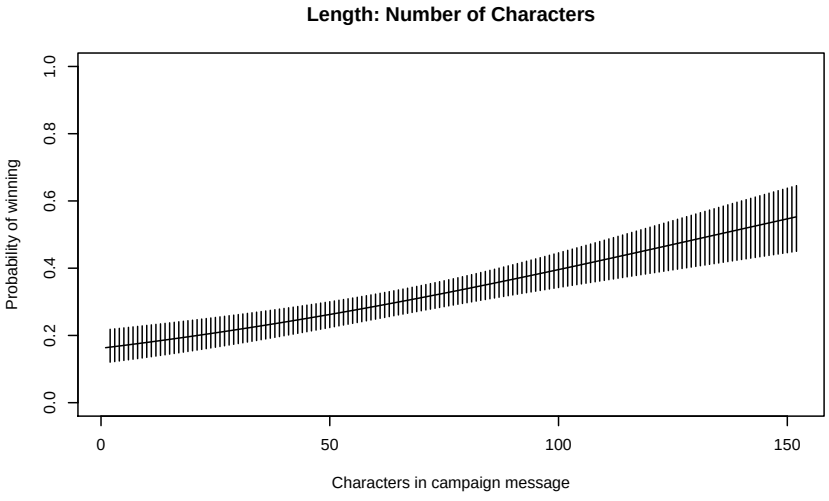
Table C.2: Effectiveness of Campaign Messages, by Category

	Dependent variable:		
	Win Election		
	(1)	(2)	(3)
	Content: Primary only	Content: All	Length
Bid High	1.601	0.617	
	(0.699)	(0.442)	
Bid Low	0.831	1.217	
	(0.754)	(0.496)	
Humor	0.737	0.104	
	(0.657)	(0.297)	
In-Team	1.119	0.265	
	(0.668)	(0.195)	
Out-Team	3.029	2.185	
	(1.366)	(1.313)	
Past Bad	1.558	0.745	
	(0.672)	(0.322)	
Skill	1.529	0.622	
	(0.628)	(0.260)	
Speed	1.865	0.359	
	(0.831)	(0.493)	
Strategy	2.543	1.454	
	(0.711)	(0.387)	
Track Record	2.419	0.970	
	(0.638)	(0.243)	
Message Length			0.012
			(0.002)
Constant	2.335	1.463	1.281
	(0.604)	(0.241)	(0.163)
Observations	638	638	638
Log Likelihood	365.942	366.270	377.254
Akaike Inf. Crit.	753.884	754.540	758.508
Note:	p<0.1; p<0.05; p<0.01		

Figure C.2: Predicted Probability of Winning Election







D Appendix: Random Selection Game Analysis

As mentioned in the manuscript and in the protocol description, participants also played a part of the game in groups with randomly selected leaders. This part of the game, which we call the Random Selection Game, took place between the individual contests (ICG) and the group election contests (DSG). The RSG was 12 rounds, like the ICG and DSG.

Comparing the ICG and DSG matches our desired comparison from a theoretical perspective. Election to political office involves both a change from representing only oneself to making choices on behalf of a group and the addition of a democratic election as selection procedure. Therefore, we believe the difference between ICG and DSG behavior - which incorporates both of these factors - captures the effect of being democratically elected. This, in turn, means that the appropriate pre-treatment behavior for our characterizations of participant type (i.e. average ticket purchases and average non-monetary value to winning) should be calculated using our data from the ICG. However, one might argue that behavior in the RSG provides an alternative measure of competitiveness. In this case, we would examine how the competitiveness of behavior as a randomly selected leader affected behavior as an elected leader.

This section reproduces the main results of the paper using the RSG instead of the ICG. Note that while we have observations for RSG behavior for all participants (see Appendix A and Appendix B for a discussion), the number of observations is far fewer than for the ICG, likely adding noise to our measures of participant type.

Overall, the analysis using the RSG as the reference stage is comparable to that of the main manuscript. More competitive women - based on RSG performance, as opposed to ICG performance - are more likely to self-select into candidacy and, to a degree, select campaign strategies that get themselves elected. This latter result is weaker than the analogous one presented in the main manuscript. However, the overall conclusion is still similar. In the main manuscript, the results suggested that self-selection into candidacy was the stronger selection channel, compared to selection effects based on the election. The results here are consistent with that argument.

Table D.1 replicates Table 1 from the main manuscript and adds the means for the RSG in the

middle row. Interestingly, women and men bought very similar numbers of tickets in the RSG, though men bought more than women in the ICG. Both genders bought more tickets as elected leaders in the DSG, compared to the RSG, and this effect was larger for women than men. The average number of tickets purchased decreased in the RSG, compared to the ICG, because we made everyone a group leader in one of the low-valued rounds. So we have a higher proportion of observations from the RSG that come from a low-valued round, compared to the ICG or DSG. Participants naturally bought fewer tickets in the low-value rounds.

Table D.1: Tickets Bought by Men and Women, by Round

Round	Overall (N)	Men (N)	Women (N)
ICG	549:8	561:5	532:6
	(1800)	(1068)	(732)
RSG	517:0	516:0	518:5
	(627)	(378)	(249)
DSG	644:2	622:8	704:3
	(240)	(177)	(63)

Table D.2 replicates Table 2 from the main manuscript. Here, we also include indicator variables for decisions made by randomly selected leaders. The first column excludes prize value fixed effects and the second column includes them. Like the main manuscript, this provides a statistical comparison of the means and differences in Table D.1. In both models, women bought significantly more tickets in the DSG than the ICG and, further, this increase was substantively larger than that between the ICG and the RSG.

Table D.3 replicates Table 3 from the main manuscript, only it uses measures of competitiveness based on behavior in the RSG, instead of that in the ICG, to compare leaders versus non-leaders, candidates versus non-candidates, and winning versus losing candidates. Again, the patterns are similar to those found using ICG-based measures of competitiveness. Looking at the top portion of the table, male leaders behaved similarly to male non-leaders, using RSG data. Male leaders actually appeared slightly less competitive in the RSG, compared to those that never led.

Table D.2: Tickets Bought by Men and Women, by Round

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Female	28.923	28.923
	(16.302)	(12.388)
RSG	45.562	4.044
	(20.333)	(15.791)
DSG	61.314	57.005
	(27.572)	(20.954)
Female RSG	31.440	47.004
	(32.167)	(24.450)
Female DSG	110.420	126.834
	(52.442)	(39.867)
Constant	561.522	119.885
	(10.396)	(13.308)
Prize FEs?	N	Y
Observations	2,667	2,667
R^2	0.011	0.430
Adjusted R^2	0.009	0.428
Residual Std. Error	339.749 (df = 2661)	258.174 (df = 2656)
F Statistic	6.043 (df = 5; 2661)	200.460 (df = 10; 2656)

Note: p<0.1; p<0.05; p<0.01. ICG and RSG ticket averages are given in 100s of tickets.

Among women, the differences between leaders and non-leaders are very stark. Women leaders purchased a great deal more tickets in the RSG and had a much higher estimated non-monetary value to winning. Looking at the middle portion of the table, the difference is similarly stark comparing women candidates and women non-candidates. Those who self-selected into candidacy purchased 115 more tickets in the average RSG round compared to those who did not. They also had much higher estimated NMVs. Looking at the bottom portion of the table, the differences are again similar though not as stark. Women who ran and won purchased more tickets in the RSG and had higher estimated NMVs, compared to women who ran and lost. These differences among women were larger than the equivalent differences for men.

Table D.3: Differences in RSG behavior, by Gender

(I) Leaders vs. Not Leaders

	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Ave. RSG Tickets	514	513	622	505
difference		+1		+117
Ave. RSG NMV	567	594	1045	613
difference		27		+432

(II) Candidates vs. Not Candidates

	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Ave. RSG Tickets	498	525	594	479
difference		27		+115
Ave. RSG NMV	517	647	934	520
difference		130		+414

(III) Winning vs. Losing Candidates

	Men		Women	
	winner	*loser*	*winner*	*loser*
Ave. RSG Tickets	502	488	615	584
difference		+14		+31
Ave. RSG NMV	519	487	1021	893
difference		+32		+128

Note: Data for "winning candidates" exclude those who ran unopposed.

Table D.4 and Table D.5 replicate Table 4 and Table 5 in the main manuscript. They also provide statistical assessment of the differences described in Table D.3. Looking at the interaction terms in Table D.4, both are positive and significant. This indicates that the competitiveness measures were more highly and positively correlated with the decision to self-select into candidacy for women than for men. Looking at the interaction terms in Table D.5, they are again positive. The competitiveness measures were more highly and positively correlated with the probability of winning for women candidates compared to men. But these differences were not as large and we cannot reject the null hypothesis that the effect of the competitiveness measures on the probability of winning was equivalent for men and women.

Table D.6 replicates Table 6 in the main manuscript. It again shows that competitiveness was associated with an increased likelihood of choosing one of the most successful campaign messages (all four interaction terms are positive). Though this relationship is more sensitive to whether we consider the primary message content or all message content. The interaction terms are only statistically significant in the latter specifications, in columns 3 and 4.

Finally, Table D.7 replicates Table 7 from the main manuscript. Looking at the interaction terms, women who were more competitive in the RSG wrote longer campaign messages (which tended to be more successful), and this effect was stronger for women than for men. Each of the interaction terms are positive and significant.

Table D.4: Effect of Competitiveness on Candidacy Decisions, by Gender

	Dependent variable:	
	Candidate	
	(1)	(2)
Female	2.473	0.958
	(0.336)	(0.130)
Ave. ICG Tickets	0.008	
	(0.031)	
Female Ave. ICG Tickets	0.352	
	(0.056)	
Ave. ICG NMV		0.002
		(0.008)
Female Ave. ICG NMV		0.088
		(0.014)
Constant	0.321	0.286
	(0.183)	(0.074)
Observations	1,786	1,786
Log Likelihood	1,141.795	1,141.795
Akaike Inf. Crit.	2,291.590	2,291.590

Note: $p<0.1$; $p<0.05$; $p<0.01$. RSG ticket averages are given in 100s of tickets.

Table D.5: Effect of Competitiveness on Electoral Success, by Gender

	Dependent variable:	
	Victory	
	(1)	(2)
Female	0.773	0.626
	(0.619)	(0.291)
Ave. RSG Tickets	0.040	
	(0.057)	
Female Ave. RSG Tickets	0.050	
	(0.103)	
Ave. RSG NMV		0.006
		(0.014)
Female Ave. RSG NMV		0.017
		(0.026)
Constant	0.893	0.728
	(0.301)	(0.126)
Observations	645	645
Log Likelihood	391.734	391.828
Akaike Inf. Crit.	791.469	791.655

Note: $p<0.1$; $p<0.05$; $p<0.01$. RSG ticket averages are given in 100s of tickets.

Table D.6: Selection of top three campaign strategies, by Gender

	Dependent variable:			
	Primary message		Part of message	
	(1)	(2)	(3)	(4)
Female	1.054	0.995	2.112	0.938
	(0.779)	(0.361)	(0.615)	(0.281)
Ave. RSG Tickets	0.057		0.069	
	(0.061)		(0.055)	
Female Ave. RSG Tickets	0.007		0.298	
	(0.127)		(0.100)	
Ave. RSG NMV		0.005		0.029
		(0.015)		(0.014)
Female Ave. RSG NMV		0.002		0.072
		(0.032)		(0.024)
Constant	1.329	1.071	0.201	0.396
	(0.329)	(0.135)	(0.291)	(0.120)
Observations	676	676	676	676
Log Likelihood	339.790	340.298	426.430	426.678
Akaike Inf. Crit.	687.581	688.597	860.861	861.356

Note: $p<0.1$; $p<0.05$; $p<0.01$. RSG ticket averages are given in 100s of tickets.

Table D.7: Length of Campaign Messages, by Effort and Gender

	Dependent variable:	
	Words in Campaign Message	
	(1)	(2)
Female	43.855	16.635
	(10.143)	(4.679)
Ave. RSG Tickets	3.874	
	(1.024)	
Female Ave. RSG Tickets	6.624	
	(1.717)	
Ave. RSG NMV		1.145
		(0.257)
Female Ave. RSG NMV		1.408
		(0.430)
Constant	85.648	72.290
	(5.433)	(2.263)
Observations	676	676
R^2	0.036	0.039
Adjusted R^2	0.032	0.035
Residual Std. Error (df = 672)	38.513	38.454
F Statistic (df = 3; 672)	8.362	9.068

Note: $p<0.1$; $p<0.05$; $p<0.01$. RSG ticket averages are given in 100s of tickets.

E Appendix: Dropping Low Value Rounds

This section of the appendix replicates all of the main manuscript's analysis, only it excludes the rounds with low prize values. In both the ICG and DSG, we included two rounds in which the prize value was lower than the ticket endowment, allowing participants to purchase more tickets than the prize was actually worth. In general, the number of tickets for every susbet of the data increases, since participants bought fewer tickets in low value rounds (LVRs). However, all of the other comparisons, especially those across gender, obtain.

Women bought fewer tickets in the ICG than men. But women leaders bought more tickets in the DSG than their male counterparts (as in Table E.1). These differences are statistically significant (as in Table E.2).

If we drop LVRs from the competitiveness measures, eg excluding LVRs from the participants' average ICG ticket purchases and estimated NMV, we again see the patterns described in the main manuscript. The difference in the competitiveness measures between leaders and non-leaders is very large for women and not as large for men (top portion of Table E.3).

The selection effects that explain why women purchased more tickets in the DSG than men operate both through self-selection into candidacy and the likelihood of winning an election. Women candidates were much more competitive than women non-candidates, and this difference is much larger for women than men (middle portion of Table E.3, statistical comparison in Table E.4). Successful women candidates were also much more competitive, and this difference was larger for women than men (bottom portion of Table E.3, statistical comparison in Table E.5.)

More competitive women also were more likely to choose successful campaign messages than less successful women, and this difference was larger for women than men (Table E.6). The same patterns were apparent in message length (Table E.7).

Table E.1: Tickets Bought by Men and Women, by Round

Round	Overall (N)	Men (N)	Women (N)
ICG	635:8	652:0	612:1
	(1500)	(890)	(610)
DSG	741:9	712:4	828:1
	(200)	(149)	(51)

Table E.2: Tickets Bought by Men and Women, by Round

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Female	39.918	39.918
	(15.309)	(14.652)
DSG	60.346	60.762
	(25.779)	(24.677)
Female DSG	155.654	154.021
	(49.666)	(47.564)
Constant	652.017	523.295
	(9.762)	(16.437)
Prize FEs?	N	Y
Observations	1,700	1,700
R^2	0.021	0.105
Adjusted R^2	0.019	0.101
Residual Std. Error	291.238 (df = 1696)	278.748 (df = 1692)
F Statistic	12.072 (df = 3; 1696)	28.417 (df = 7; 1692)

Note: p<0.1; p<0.05; p<0.01.

Table E.3: Differences in ICG behavior, by Gender

(I) Leaders vs. Not Leaders

	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Ave. ICG Tickets	565	557	670	520
difference		+8		+150
Ave. ICG NMV	539	507	958	360
difference		+32		+598

(II) Candidates vs. Not Candidates

	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Ave. ICG Tickets	560	557	616	495
difference		+3		+121
Ave. ICG NMV	520	506	741	259
difference		+14		+482

(III) Winning vs. Losing Candidates

	Men		Women	
	winner	*loser*	*winner*	*loser*
Ave. ICG Tickets	561	557	675	595
difference		+4		+80
Ave. ICG NMV	522	598	979	662
difference		76		+317

Note: Data for "winning candidates" exclude those who ran unopposed.

Table E.4: Effect of Competitiveness on Candidacy Decision, by Gender

	Dependent variable:	
	Candidate	
	(1)	(2)
Female	2.473	0.958
	(0.336)	(0.130)
Ave. ICG Tickets	0.008	
	(0.031)	
Female Ave. ICG Tickets	0.352	
	(0.056)	
Ave. ICG NMV		0.002
		(0.008)
Female Ave. ICG NMV		0.088
		(0.014)
Constant	0.321	0.286
	(0.183)	(0.074)
Observations	1,786	1,786
Log Likelihood	1,141.795	1,141.795
Akaike Inf. Crit.	2,291.590	2,291.590

Note: $p<0.1$; $p<0.05$; $p<0.01$. ICG ticket averages are given in 100s of tickets.

Table E.5: Effect of Competitiveness on Electoral Success, by Gender

	Dependent variable:	
	Victory	
	(1)	(2)
Female	1.991	0.940
	(0.686)	(0.288)
Ave. ICG Tickets	0.008	
	(0.049)	
Female Ave. ICG Tickets	0.244	
	(0.104)	
Ave. ICG NMV		0.002
		(0.012)
Female Ave. ICG NMV		0.061
		(0.026)
Constant	0.739	0.706
	(0.291)	(0.121)
Observations	645	645
Log Likelihood	388.531	388.531
Akaike Inf. Crit.	785.062	785.062

Note: $p<0.1$; $p<0.05$; $p<0.01$. ICG ticket averages are given in 100s of tickets.

Table E.6: Selection of Top Three Campaign Strategies, by Gender

	Dependent variable:			
	Primary message		Part of message	
	(1)	(2)	(3)	(4)
Female	2.612	1.425	2.420	0.882
	(0.847)	(0.354)	(0.614)	(0.250)
Ave. ICG Tickets	0.115		0.239	
	(0.052)		(0.050)	
Female Ave. ICG Tickets	0.276		0.357	
	(0.128)		(0.097)	
Ave. ICG NMV		0.029		0.060
		(0.013)		(0.012)
Female Ave. ICG NMV		0.069		0.089
		(0.032)		(0.024)
Constant	0.410	0.905	0.770	0.259
	(0.298)	(0.122)	(0.286)	(0.114)
Observations	676	676	676	676
Log Likelihood	336.910	336.910	417.818	417.818
Akaike Inf. Crit.	681.821	681.821	843.637	843.637

Note: p<0.1; p<0.05; p<0.01. ICG ticket averages are given in 100s of tickets.

Table E.7: Length of Campaign Messages, by Effort and Gender

	Dependent variable:	
	Words in Campaign Message	
	(1)	(2)
Female	38.598	16.420
	(10.531)	(4.261)
Ave. ICG Tickets	2.345	
	(0.876)	
Female Ave. ICG Tickets	5.154	
	(1.673)	
Ave. ICG NMV		0.586
		(0.219)
Female Ave. ICG NMV		1.288
		(0.418)
Constant	79.443	69.352
	(5.248)	(2.159)
Observations	676	676
R^2	0.026	0.026
Adjusted R^2	0.021	0.021
Residual Std. Error (df = 672)	38.717	38.717
F Statistic (df = 3; 672)	5.919	5.919

Note: $p<0.1$; $p<0.05$; $p<0.01$. ICG ticket averages are given in 100s of tickets.

F Appendix: Alternative Explanations

In this section of the appendix, we consider several alternative explanations for our findings. The DSG portion of the game adds an election, which potentially induces two changes that are distinct from selection effects. First, the experience of being elected can affect behavior directly as established in (citation removed for anonymity). This "election effect" could vary across gender. We consider this first and do not find any evidence that the election effect varies by gender.

Second, there could be re-election concerns in the DSG, since groups are only shuffled every other round. We think this is unlikely, since participants are always unidentified and anonymous and therefore don't know who the incumbent is. Nevertheless, we investigate whether there is evidence of re-election concerns for men or women. We do not find evidence of this concern for either subset or in the full set of participants.

In the third subsection of this appendix, we provide summary information about the distribution of candidacy decisions and election outcomes, broken down by gender. We then replicate the main analyses excluding respondents who were outliers in terms of how frequently they were leader.

F.1 Election Effects

In this section, we rule out the possibility that our results are driven by a differential election effect in which the experience of being elected affects women more than men. To do this, we regress the number of tickets purchased on indicators for the DSG part of the game. We include participant fixed effects, so that we are estimating within-participant changes in ticket purchases across different parts of the game. As above, we include prize fixed effects to account for differing values across rounds. We split the sample between men and women for ease of interpretation.

Table F.1 shows that the election effect was, in fact, somewhat more prominent among men than women. The ICG is the base category. Among men, elected leaders in the DSG purchased approximately 54 more tickets, compared to their average ticket purchases. Women leaders purchased approximately 53 more tickets. The election effect matters for both men and women, but it

is larger for men. We therefore conclude that the differences in ticket purchasing behavior among elected men and women is not driven by a differential election effect.

Table F.1: Within-Participant Changes in Tickets Bought for Eventual Leaders, by Gender

	Dependent variable:		
	Tickets Bought		
	Sample:		
	Men Only	Women Only	Full
	(1)	(2)	(3)
DSG	53.983	52.994	54.538
	(16.473)	(24.738)	(16.176)
DSG Female			1.555
			(30.618)
Constant	121.146	74.888	
	(46.056)	(45.307)	
Prize FEs?	Y	Y	Y
Participant FEs?	Y	Y	Y
Observations	1,011	514	1,525
R^2	0.709	0.721	0.925
Adjusted R^2	0.692	0.702	0.920
Residual Std. Error	192.322 (df = 954)	179.479 (df = 480)	189.022 (df = 1439)
F Statistic	41.484 (df = 56; 954)	37.569 (df = 33; 480)	205.674 (df = 86; 1439)

Note: p<0.1; p<0.05; p<0.01.

F.2 Re-election Effects

Table F.2 first assesses whether there is any re-election effect in the DSG. We regress the number of tickets purchased in the DSG on an indicator variable for odd numbered rounds (Reelection round), since groups are reshuffled after even numbered rounds. Ticket purchases were not higher in these rounds. Column 2 adds an indicator variable for women respondents and the interaction

between that and the reelection round variable. Women purchased approximately 13 more tickets in these rounds compared to men, but this difference is not substantively or statistically significant.

Table F.2: Reelection effects, by Gender

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Reelection round	9.803	15.292
	(12.480)	(14.915)
Female		20.026
		(14.962)
Reelection Female		13.583
		(20.657)
Prize FEs?	Y	Y
Observations	2,667	2,667
R^2	0.421	0.422
Adjusted R^2	0.420	0.420
Residual Std. Error	259.988 (df = 2660)	259.988 (df = 2658)
F Statistic	322.628 (df = 6; 2660)	242.220 (df = 8; 2658)

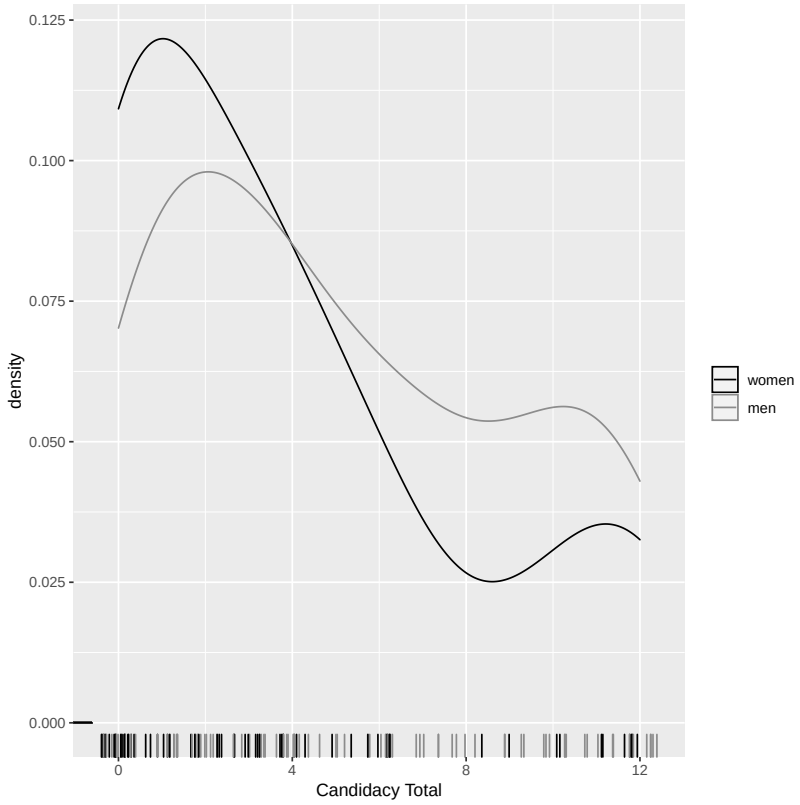
Note: p<0.1; p<0.05; p<0.01.

F.3 Candidacy/Election by Gender, Analysis Excluding Outliers

As mentioned in the main manuscript, men more frequently ran for and won the elections. The percentage of women who ran for elections at least once was 72% while the percentage for men was about 88%. Women were candidates 31% of the time while men were candidates 43% of the time. Moreover, 46% of women were leaders at least once while 57% of men were leaders at least once. Women were leaders in 26.25% of the DSG rounds while men were leaders in 73.75% of the rounds. Figure F.1 shows the smoothed distribution of the number of times a participant ran for election. Women candidates were more likely to never run for office, and male candidates

were more likely to always run for office. Figure F.2 shows similar patterns when looking at the distribution of the number of times a participant was leader, by gender. Women were much more likely than men to never be leader. The four participants who were leader most frequently were all male.

Figure F.1: Distribution of number of times running for election, by Gender

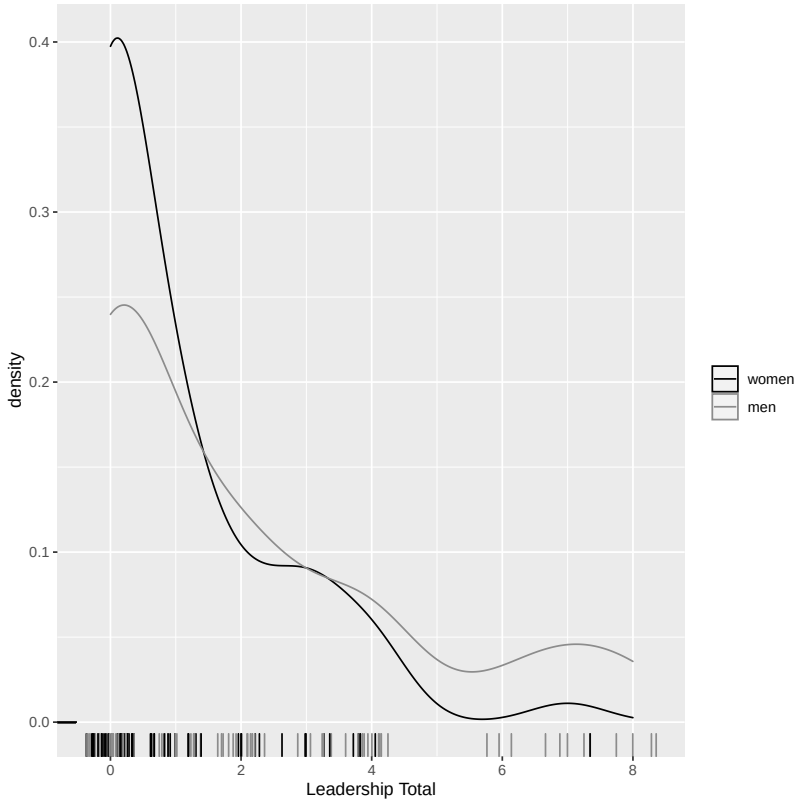


Note: This shows the smoothed distribution for the number of times each participant ran for election. The tick marks in along the horizontal axis are clustered around discrete values, ranging from 0-12.

One woman was leader in 7 rounds. The next highest total for women participants was 4 rounds. The woman participant who was leader in 7 rounds was also especially competitive. She purchased a very high number of tickets in the ICG. It is therefore worth checking whether this outlier participant is driving our results.

In short, the answer is no. Table F.3 through Table F.5 replicate Tables 2, 4, and 5 from the main manuscript. The differences between men and women leaders shrink somewhat, per Table F.3, but

Figure F.2: Distribution of number of times winning election, by Gender



Note: This shows the smoothed distribution for the number of times each participant won an election. The tick marks in along the horizontal axis are clustered around discrete values, ranging from 0-8.

they are still significant and meaningful. The same is true when comparing how competitiveness measures affect candidacy decisions across men and women, per Table F.4. The main difference when excluding this participant is in Table F.5. There, competitiveness did not have a significantly different effect on the likelihood of winning, conditional on candidacy, for men and women. This is consistent with the main manuscript's emphasis on self-selection as being more important than group selection.

Excluding the male outliers who were frequently leaders actually strengthens the size and significance of the differences across gender highlighted in the main manuscript. The four men who were frequently leaders actually had slightly higher ticket purchases in the ICG, compared to the overall average for male leaders. This means that excluding them *increases* the differences between

Table F.3: Tickets Bought by Men and Women, by Round, Excluding Outlier Woman Leader

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Female	35.463	35.463
	(16.343)	(12.813)
DSG	61.314	57.228
	(27.506)	(21.567)
Female x DSG	96.073	112.302
	(54.474)	(42.730)
Constant	561.522	124.963
	(10.371)	(15.516)
Prize FEs?	N	Y
Observations	2,021	2,021
R^2	0.010	0.393
Adjusted R^2	0.009	0.391
Residual Std. Error	338.929 (df = 2017)	265.709 (df = 2012)
F Statistic	6.885 (df = 3; 2017)	162.925 (df = 8; 2012)
Note:		p<0.1; p<0.05; p<0.01

Table F.4: Effect of Competitiveness on Candidacy Decision, by Gender, Excluding Outlier Woman Leader

	Dependent variable:	
	Candidate	
	(1)	(2)
Female	2.294	0.938
	(0.338)	(0.129)
Ave. ICG Tickets	0.008	
	(0.031)	
Female Ave. ICG Tickets	0.315	
	(0.057)	
Ave. ICG NMV		0.002
		(0.008)
Female Ave. ICG NMV		0.079
		(0.014)
Constant	0.321	0.286
	(0.183)	(0.074)
Observations	1,774	1,774
Log Likelihood	1,135.921	1,135.921
Akaike Inf. Crit.	2,279.842	2,279.842
Note:	p<0.1; p<0.05; p<0.01	
	Averages given in 100s of tickets.	

Table F.5: Effect of Competitiveness on Electoral Success, by Gender, Excluding Outlier Woman Leader

	Dependent variable:	
	Victory	
	(1)	(2)
Female	1.634	0.868
	(0.716)	(0.286)
Ave. ICG Tickets	0.008	
	(0.049)	
Female Ave. ICG Tickets	0.178	
	(0.112)	
Ave. ICG NMV		0.002
		(0.012)
Female Ave. ICG NMV		0.045
		(0.028)
Constant	0.739	0.706
	(0.291)	(0.121)
Observations	633	633
Log Likelihood	379.294	379.294
Akaike Inf. Crit.	766.588	766.588
Note:	$p<0.1$; $p<0.05$; $p<0.01$	
	Averages given in 100s of tickets.	

men and women leaders, in general, since women leaders tended to be more competitive than male leaders in the full sample. Table F.6 through Table F.8 show this, again replicating Tables 2, 4, and 4 from the main manuscript.

Table F.6: Tickets Bought by Men and Women, by Round, Excluding Male Outliers

	Dependent variable:	
	Tickets Bought	
	(1)	(2)
Female	27.646	27.646
	(16.470)	(13.057)
DSG	57.133	50.503
	(30.176)	(23.932)
Female x DSG	114.600	132.404
	(53.883)	(42.752)
Constant	560.246	128.091
	(10.646)	(16.028)
Prize FEs?	N	Y
Observations	1,960	1,960
R^2	0.010	0.380
Adjusted R^2	0.009	0.377
Residual Std. Error	339.999 (df = 1956)	269.539 (df = 1951)
F Statistic	6.737 (df = 3; 1956)	149.180 (df = 8; 1951)
Note:		$p<0.1$; $p<0.05$; $p<0.01$

Table F.7: Effect of Competitiveness on Candidacy Decision, by Gender, Excluding Male Outliers

	Dependent variable:	
	Candidate	
	(1)	(2)
Female	2.394	0.870
	(0.338)	(0.131)
Ave. ICG Tickets	0.006	
	(0.031)	
Female Ave. ICG Tickets	0.354	
	(0.056)	
Ave. ICG NMV		0.002
		(0.008)
Female Ave. ICG NMV		0.089
		(0.014)
Constant	0.400	0.374
	(0.186)	(0.075)
Observations	1,738	1,738
Log Likelihood	1,101.855	1,101.855
Akaike Inf. Crit.	2,211.710	2,211.710
Note:	$p<0.1$; $p<0.05$; $p<0.01$	
	Averages given in 100s of tickets.	

Table F.8: Effect of Competitiveness on Electoral Success, by Gender, Excluding Male Outliers

	Dependent variable:	
	Victory	
	(1)	(2)
Female	1.864	0.817
	(0.692)	(0.291)
Ave. ICG Tickets	0.008	
	(0.051)	
Female x Ave. ICG Tickets	0.243	
	(0.105)	
Ave. ICG NMV		0.002
		(0.013)
Female x Ave. ICG NMV		0.061
		(0.026)
Constant	0.866	0.830
	(0.304)	(0.128)
Observations	610	610
Log Likelihood	358.422	358.422
Akaike Inf. Crit.	724.843	724.843
Note:	$p<0.1$; $p<0.05$; $p<0.01$	
	Averages given in 100s of tickets.	

G Appendix: Competitiveness vs. Con dence

In the main text, we showed that women who purchased more tickets in the Individual Contest Games were more likely to self-select into and win democratic elections. We associated this with competitiveness, but a related concept, con dence, also potentially affects selection into candidacy. In this section of the appendix, we examine the effect of con dence by looking at whether obtaining greater payoffs in the Individual Contest Games (ICG) affected the gender gap in self-selection as well as in group selection. We expect that earning high payoffs in ICG contest games would boost self-con dence in one's own ability to play the game well.

Existing research focuses on different aspects of con dence. First, a participant may be more con dent in her objective, underlying ability to complete a task. For example, in Kanthak and Woon (2015), participants add two digit numbers. A participant's ability to correctly and quickly add numbers is a measurable trait. Knowing one's pro ciency can boost (or decrease) a participant's con dence. Second, participants also con dence in their abilities relative to others. This aspect of con dence speaks to beliefs about where one's abilities fall in the distribution of participant abilities. Existing research usually measures these beliefs by asking an incentivized question about how a participant's abilities compare to others'.

Unlike much existing work, the underlying task in our game is not one where participants have an “objective” skill level. The underlying interaction is strategic; a participant's payoff depends on her choice and that of her opponent. Additionally, payoffs have a stochastic element. A participant might choose a strategy that maximizes her expected utility, but nonetheless lose a contest.

However, we can use a participant's average payoff in the ICG as a proxy for con dence. Having a higher payoff in the one-on-one contests relates to beliefs about one's ability. If a participant does better in the ICG, she has concrete information about the utility she accrues from her choices. She also gets information about her performance relative to others, since we showed each pair the outcomes of the contest, ticket purchases, and payoffs, after every round.

This measure cannot distinguish between con dence stemming from a participant's ability to choose strategies that maximize her expected utility and a participant's beliefs about her own luck.

However, a participant's ICG payoff is an appropriate measure for con dence in our setting because it is shaped by the information each participant directly receives in the game. For our participants, their payoffs in the ICG are likely their best pieces of information about their ability to achieve higher payoffs.

A third aspect of con dence refers to one's beliefs about their ability to win elections. This refers to beliefs about how a participant would fare in an election, separate from her ability to do well at an underlying task. This facet of con dence is also inherently relational, since belief in one's ability to win elections implies a belief that one can win an election relative to other possible candidates. Pruysers and Blais (2017) and Fox and Lawless (2011) link this aspect of con dence with gender differences in election aversion. Though, Bernhard and de Benedictis-Kessner (2020) do not nd that women candidates are especially discouraged by electoral losses.

Con dence as measured by ICG payoffs

To assess the relationship between con dence and choices in our game, Table G.1 replicates a table from the main manuscript, showing average ICG payoffs for leaders versus non-leaders. This table is constructed in the same way as Table 3 which showed how leaders and non-leaders differed by our competitiveness measures.

Leaders in the DSG tended to have higher ICG payoffs than non-leaders (top panel). However, these differences are not especially large. And more importantly, they are similar for men and women. The same is true when looking at the ICG payoffs of candidates versus non-candidates (middle panel) and winning versus losing candidates (bottom panel).

Table G.2 shows these relationships by regressing an indicator variable for whether a participant self-selected into candidacy on their average ICG payoff, with a gender interaction term (Column 1). Looking at the interaction term, average ICG payoff matters slightly less for women's decisions to select into candidacy, but the effect is very close to zero and insigni cant.

In Column 2, we also include our measures of competitiveness based on average ICG ticket purchases, and in Column 3, we do the same for our measure of competitiveness based on estimates

Table G.1: Differences in ICG payoff, by Gender

(I) Leaders vs. Not Leaders

	Men		Women	
	leader	not leader	leader	not leader
Ave. ICG Payoff	1401	1308	1377	1310
difference	+93		+67	

(II) Candidates vs. Not Candidates

	Men		Women	
	candidate	not candidate	candidate	not candidate
Ave. ICG Payoff	1346	1306	1328	1311
difference	+40		+17	

(III) Winning vs. Losing Candidates

	Men		Women	
	winner	loser	winner	loser
Ave. ICG Payoff	1388	1313	1369	1310
difference	+75		+59	

Note: Data for “winning candidates” exclude those who ran unopposed.

of NMV. In both specifications, the confidence measure based on ICG payoffs has no different effect for men versus women. Yet, even including this measure of confidence, our measures of competitiveness still have similar, significantly different effects for men versus women, as in the main manuscript. Controlling for our proxy for confidence, the significance of the average ICG tickets purchased, our measure for competitiveness, persists.

Table G.3 replicates that analysis again focusing on the likelihood of winning an election, conditional on selecting into candidacy. The dependent variable is a binary indicator for whether the candidate wins. As above, we cannot reject the null that average ICG payoff has the same effect for men as for women. And as above, our measures of competitiveness still do have different effects on the probability of winning for men and women candidates.

Additionally, for confidence in abilities to explain our results, it would need to be the case that boosts to confidence - stemming from a higher payoff in one round of the ICG - would need to be associated with a higher ticket purchase in ensuing rounds, and this effect would need to differ by gender. We see neither relationship. Table G.4 shows a regression of ICG ticket purchases on participants' lagged payoffs from the previous round, interacted with gender. Column 1 excludes prize level fixed effects; Column 2 includes them. A higher payoff in round t tends to be associated with slightly lower purchases in round $t+1$ and this effect is not different for men versus women in either specification.

Table G.2: Effect of ICG Payoff on Candidacy, by Gender

	Dependent variable:		
	Candidate		
	(1)	(2)	(3)
Female	0.013	2.690	1.126
	(0.489)	(0.656)	(0.527)
Ave. ICG Payoff	0.001	0.001	0.001
	(0.0002)	(0.0002)	(0.0002)
Ave. ICG Tickets		0.022	
		(0.031)	
Ave. ICG NMV			0.006
			(0.008)
Female Ave. ICG Payoff	0.0004	0.0001	0.0001
	(0.0004)	(0.0004)	(0.0004)
Female Ave. ICG Tickets		0.364	
		(0.058)	
Female ICG NMV Ave.			0.091
			(0.014)
Constant	1.039	1.204	1.108
	(0.320)	(0.395)	(0.334)
Observations	1,786	1,786	1,786
Log Likelihood	1,171.483	1,135.313	1,135.313
Akaike Inf. Crit.	2,350.966	2,282.625	2,282.625

Note: $p<0.1$; $p<0.05$; $p<0.01$
Averages given in 100s of tickets.

Table G.3: Effect of ICG Payoffs on Electoral Success, by Gender

	Dependent variable: Winning Candidate		
	(1)	(2)	(3)
Female	0.230	2.145	1.091
	(0.963)	(1.319)	(1.066)
Ave. ICG Payoff	0.001	0.001	0.001
	(0.0004)	(0.0004)	(0.0004)
Ave. ICG Tickets		0.036	
		(0.052)	
Ave. ICG NMV			0.009
			(0.013)
Female Ave. ICG Payoff	0.0001	0.0001	0.0001
	(0.001)	(0.001)	(0.001)
Female Ave. ICG Tickets		0.245	
		(0.108)	
Female ICG NMV Ave.			0.061
			(0.027)
Constant	1.975	2.254	2.097
	(0.519)	(0.655)	(0.548)
Observations	645	645	645
Log Likelihood	388.278	383.319	383.319
Akaike Inf. Crit.	784.557	778.639	778.639

Note: $p < 0.1$; $p < 0.05$; $p < 0.01$
Averages given in 100s of tickets.

Table G.4: Effect of Lagged Payoffs on Ticket Purchases, by Gender

	Dependent variable:	
	(1)	(2)
Female	30.088	39.383
	(29.004)	(22.494)
Payoff (lagged)	0.051	0.029
	(0.011)	(0.009)
Female x Payoff (lagged)	0.002	0.008
	(0.018)	(0.014)
Observations	1,650	1,650
R^2	0.022	0.414
Prize FE	No	Yes
Adjusted R^2	0.020	0.411
Residual Std. Error	343.144 (df = 1646)	265.989 (df = 1641)
F Statistic	12.216 (df = 3; 1646)	144.922 (df = 8; 1641)
Note:		$p<0.1$; $p<0.05$; $p<0.01$

Con dence as measured by electoral wins

Here, we return to the aspect of con dence that is speci c to one's ability to win elections. Women and men de nitely differ in these beliefs, as evidenced by gender differences in the likelihood of selecting into candidacy and winning elections. However, this difference alone is insuf cient to explain our results in the main manuscript. Even if men and women differ in their overall levels of electoral con dence, this would not explain why more competitive women select into and win elections. For differences in con dence in electoral ability to explain our results, boosts to electoral con dence - from winning an election - would have to affect women and men differently.

We do not nd evidence of this. Table G.5, Column 1, shows the results from regressing a binary indicator for whether a participant chose to be a candidate in round t on indicator variables for whether that participant was a winning or losing candidate in round $t-1$.[49] We also interact these indicators with gender. Candidacy decisions tend to be persistent. Both men and women

[49]Note, this analysis drops round 1 of the DSG by construction.

who were winning and losing candidates in round t are more likely to run again in round $t+1$. However, the effect of being a winning candidate does not differ by gender. The effect of being a losing candidate does differ by gender, but in the opposite way as we would expect if men and women reacted to boosts (or in this case hits) to their electoral con dence. Women candidates who lost were more persistent in their candidacy decisions than men. If anything, this suggests that the type of women who run are less in uenced by in-game con dence effects than the type of men who run. In other words, just as winning an election isn't disproportionately affecting women, losing an election is not disproportionately discouraging women.

Column 2 replicates that analysis but uses a binary indicator for whether a participant won election, conditional on candidacy. Here, too, we do not nd the con dence boosts or hits affect men versus women differently. Winning or losing in round t did not have a differential effect on the likelihood of winning in round $t+1$ for women compared to men.

Table G.5: Effect of Electoral Wins/Losses on Candidacy/Winning, by Gender

	Dependent variable:	
	Candidate	Electoral Success
	(1)	(2)
Female	0.554	0.237
	(0.165)	(0.350)
Winning Candidate (lagged)	2.205	0.821
	(0.208)	(0.284)
Losing Candidate (lagged)	1.598	0.025
	(0.161)	(0.267)
Female x Winner (lagged)	0.321	0.394
	(0.375)	(0.528)
Female x Loser (Lagged)	0.599	0.102
	(0.265)	(0.464)
Observations	1,636	571
Log Likelihood	898.695	344.372
Akaike Inf. Crit.	1,809.391	700.745
Note:	$p<0.1$;	$p<0.05$; $p<0.01$

Appendix References

Anderson, Jon, Stephen Burks, Jeffrey Carpenter, Lorenz Gotte, Karsten Maurer, Daniele Nosenzo, Ruth Potter, Kim Rocha and Aldo Rustichini. 2013. "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples." Experimental Economics 16(2):170–189.

Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon.com's mechanical turk." Political Analysis 20(3):351–368.

Bernhard, Rachel and Justin de Benedictis-Kessner. 2020. "The Boys—and Girls—Are Back in Town: Male and Female Candidates Are Similarly Dissuaded by Losing Elections.".

Fox, Richard and Jennifer 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.

Hendriks, Achim. 2012. "SoPHIE-Software platform for human interaction experiments." University of Osnabrueck, working paper

Kanthak, Kristin and Jon Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." American Journal of Political Science 59(3):595–612.

Mullinix, Kevin, Thomas Leeper, James Druckman and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." Journal of Experimental Political Science 2(2):109–138.

Pruysers, Scott and Julie Blais. 2017. "Why Won't Lola Run? An Experiment Examining Stereotype Threat and Political Ambition." Politics & Gender 13(2).