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

Yon Soo Park, Sarah Hummel, and Stephen Chaudoin

April 21, 2023*

Abstract

Most arguments for why women leaders might pursue aggressive policies during interstate conflict emphasize gendered stereotyping of international politics. We highlight a different explanation based on the self selection of certain women into candidacy for leadership positions. To illustrate this mechanism, we conduct a laboratory experiment using online real-time, group play where participants choose to run for election, run a simple campaign, and represent their group in a contest game if elected. We find that women who place a higher non-monetary value on winning were more likely to select into candidacy, win election, and then spend more resources in intergroup contests than their male counterparts. As a result, female leaders pursued more aggressive strategies than male leaders, even though women, on average, tended to invest less in one-on-one contests. These patterns appear even though our protocol stacks the deck against finding gender differences by anonymizing participants and shuffling groups. Our findings emphasize the agency and preferences of female leaders who choose to run and, subsequently, choose to fight harder in intergroup contests.

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

"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, [you] would never have got through."

– Margaret Thatcher, quoted by Gail Sheehy, *Vanity Fair* 1989

1 Introduction

The relationship between a leader's gender and interstate conflict is complicated, both theoretically and empirically. Several empirical analyses find that female chief executives pursue more hawkish foreign policies than male chief executives, despite the fact that women in the general public espouse more pacifist opinions.[1] Though, this finding about leader gender is not universal[2] and comes with caveats about the contributing role of foreign leaders' biases[3] or the moderating role of societal and institutional factors.[4]. Authors most often explain this divergence by highlighting the gendered nature of the strategic environment. Female politicians face additional pressure to combat stereotypes and prove their toughness, especially in the domains of interstate conflict and national security. Faced with the prevailing perception that women are weak, female leaders are forced to compensate by signaling their strength or resolve.

We do not doubt the gendered strategic environment influences decisions made by female executives in the international arena. However, these explanations de-emphasize preference heterogeneity across women. Any woman, faced with the realities of the gendered political environment, is pushed to behave a certain way.

In this research note, we develop a theoretical framework that directly incorporates variation in the preferences of women. Drawing on research about the gendered aspects of domestic demo-

[1]Caprioli and Boyer (2001); Dube and Harish (2020); Koch and Fulton (2011); Powell and Mukazhanova-Powell (2019); Schramm and Stark (2020*a*); Trager and Barnhart (2023).

[2]Horowitz, Stam and Ellis (2015); Imamverdiyeva and Shea (2022).

[3]Post and Sen (2020); Reiter and Wolford (2022).

[4]Eg Koch and Fulton (2011); Schramm and Stark (2020*a*).

cratic selection, we argue that hawkish female leaders self-select into candidacy for leadership positions at higher rates than other women. Existing research documents a greater competition aversion among women, which makes the prospect of engaging in electoral contests intrinsically less attractive. Furthermore, the expected benefits of being in office may be lower for women, precisely because of the gender biases they expect to face in office. In short, running for office is costlier for women than men. As a result, only women with a strong intrinsic drive to win contests - a high "non-monetary value to winning" - choose to enter elections. If they succeed, this same drive affects their choices in inter-group conflicts, leading them to pursue more aggressive strategies in the pursuit of victory.

Identifying the effect of leader gender in observational studies is difficult because leader gender is not randomly assigned.[5] As Cohen and Karim (2021) describe in their call for more micro-level evidence linking gender and conflict, endogeneity concerns make it difficult to analyze theoretical mechanisms because of non-random selection. Selection mechanisms are especially difficult to isolate in observational studies because doing so requires measuring key features of people who do *and do not* eventually become leaders, from a pool of possible leaders. The list of female leaders is also frustratingly short, which raises the possibility that specific leaders or situations have out-sized influence on estimates.[6]

To address these concerns, we turn to a controlled laboratory setting. Our participants played contest games where they chose how much costly effort to expend to try and win a zero sum prize. They first played as individuals in one-on-one contests. They then played in groups that endogenously elected leaders to represent their group. Participants decided whether to run for election, wrote short campaign messages, and voted for their preferred candidates. Participants were anonymous throughout the game and groups were shuffled frequently, which decreases the role that stereotyping can play in the election.

Our results first reproduce the observation motivating many studies of leader gender and conflict. Female elected leaders pursued more aggressive strategies in intergroup contests compared to

[5] Schwartz and Blair (2020).

[6] Imamverdiyeva and Shea (2022).

male leaders, despite that fact that women, on average, chose less aggressive strategies in individual contests. These differences between male and female leaders arise through non-random selection via the election procedure. Consistent with our theoretical expectations, women with higher non-monetary values to winning contests were significantly more likely to run for office and win. This dynamic was much less stark for men. By analyzing campaigns, we show that women with higher non-monetary values also exerted more effort in their campaigns by writing longer and more effective campaign messages. We additionally show that selection occurs based on non-monetary values to winning rather than risk-aversion or different levels of participant confidence. The correlation between measures of risk aversion and confidence and selection into leadership are not strong for women, nor do they differ substantially from analogous correlations among men.

Our study thus highlights a mechanism, further up the "pipeline" than many others, to help explain why women leaders might pursue more aggressive strategies in conflict situations. We are not the first to consider the role of selection. For instance, Enloe (1989) wondered whether female leaders are outliers with respect to conflict behavior who - in Margaret Thatcher's words - "got through."[7] In a call for future research over 25 years later, Reiter (2015) likewise posits that the "conflict attitudes of women elected to office [may not] represent the conflict attitudes of all women" (1313). Our study design lets us answer those calls and provide direct evidence supporting the importance of self-selection. By specifying the traits driving selection and leader choices, it also develops a more complete theoretical framework for understanding how the process of selection can result in an over-representation of female leaders who are unafraid of conflict.

To reiterate, our goal is not to suggest that selection matters more than the strategic environment for explaining the choices of women leaders. The two are not mutually exclusive. But our research does suggest that gendered patterns of self-selection into leadership affect the observed differences in behavior between male and female leaders.[8] The stark selection effects we find suggest that estimates of the effect of gender on leader choices from observational data are potentially biased.

[7] Interview with Gail Sheehy, *Vanity Fair* 1989.

[8] We recognize the distinction between sex and gender. We follow the literature in using the term gender to describe leader attributes and their strategic environment.

By emphasizing the preferences of women who become leaders, our explanation - as Judge (2021) writes - focuses on "the interests and calculations of female leaders themselves," as opposed to only "the actors who create the structural space for a woman to attain leadership" (4).

We acknowledge the caution needed in drawing inferences about leaders and conflict from laboratory studies. In this case, however, the advantages of a laboratory study outweigh the disadvantages. Not only are we able to measure key attributes of both leaders and non-leaders, but we can also minimize gendered aspects of the strategic environment during conflict. This is not possible in the real world since the gender of leaders is necessarily known. We would also note that recent analysis of over 100 paired experiments on political elites and mass publics found minimal gaps between the two groups.[9] Following Renshon, Lee and Tingley (2017), we view the evidence presented here as "part of larger bodies of evidence compiled from different data sources and different research designs" (S204).

2 Gender and Self-Selection

The relationship between gender and conflict is complex. On the one hand, survey responses and survey experimental data consistently find that female citizens have more pacifist preferences. 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, which persists cross-nationally. Meta-analysis of 17 recent survey experiments also finds that female respondents are less supportive of the use of force.[10]

Yet, research focusing on female *leaders* often finds that they choose equally or more hawkish policies than their male counterparts. Existing research finds that female leaders are involved in more militarized interstate disputes, international conflicts, or have higher levels of defense spending.[11] Horowitz, Stam and Ellis (2015) find a positive relationship between female leaders

[9] Kertzer (Forthcoming).

[10] Barnhart et al. (2020).

[11] Burns and Bowling (2021); Caprioli and Boyer (2001); Dube and Harish (2020); Koch and Fulton (2011); Powell and Mukazhanova-Powell (2019); Schramm and Stark (2020*b*); Trager and Barnhart (2023).

and MID involvement, though this finding is insignificant, which they attribute to the paucity of female leaders in the data. Imamverdiyeva and Shea (2022) find that differences between male and female leaders' conflict patterns are heavily influenced by statistical outliers, like Golda Meir. Some researchers find that general levels of societal gender equality moderate the relationship between leader gender and conflict,[12] while others emphasize that the effect of leader gender is tempered by the number of female legislators,[13] or exacerbated by the presence of higher executive constraints.[14] Swers (2007) finds that female members of Congress are more likely to sponsor defense bills, though, given the relatively small number of "hard" defense bills, she emphasizes this pattern for homeland security bills and "soft" defense bills.

The most prominent explanations for the disconnect between conflict preferences of female citizens and the choices of female executives emphasize gender bias. Negative stereotypes of women as weak force female leaders to compensate with policies that counteract that stereotype. Pressure can come from voters, whose biases push women to "prove themselves" by taking actions that run counter to these stereotypes.[15] Blair and Schwartz (2021) find that respondents are less approving of conciliatory policies from female leaders. Post and Sen (2020) argue that stereotyping from opposing leaders also affect conflict escalation. Female leaders are perceived as less resolved, so when they are the challenger in an interstate dispute, the target is more likely to reciprocate the dispute and the female leader will be more likely to escalate the dispute to higher hostility levels.

Stereotypes also complicate predictions relating gender to crisis bargaining. Reiter and Wolford (2022) model sexist leaders who may gain a private benefit from defeating a female-led adversary or suffer private costs to losing to one. The effect of having a female leader on the probability of war depends on the probability that her (potentially sexist) opponent will win a conflict. Disputes against and war with a female-led adversary are more likely for confident sexist leaders, but less likely for those with lower probabilities of winning disputes. Schwartz and Blair (2020) show that respondents in a survey experiment more harshly punish female leaders for backing down from

[12] E.g. Caprioli and Boyer (2001); Powell and Mukazhanova-Powell (2019).
[13] Koch and Fulton (2011).
[14] Schramm and Stark (2020*b*).
[15] E.g. Lawless (2004).

threats. This, too, has conflicting implications for the relationship between leader gender and war. Greater audience costs could facilitate stronger signals of resolve from female leaders which might have a pacifying effect, but they could also encourage adventurism and make it harder for a female leader to de-escalate after a threat (p 890).

We highlight an additional explanation: the leader selection process means that women rising to the role of chief executive are not representative of the broader population. Rather, they place a higher value on winning from the start and therefore choose more aggressive policies once elected. While not mutually exclusive with explanations based on gender stereotyping, the non-randomness of democratic selection is an important channel through which gender affects conflict.

The initial step in our theory draws on Niederle and Vesterlund (2007) and a large body of subsequent literature documenting competition aversion among women.[16] In experimental economic work, numerous studies establish that women experience greater competition aversion than men 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 competition.[17] Elections are zero-sum contests, since only one candidate can win. While several studies limited gender bias in the election rates of comparable men and women, 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 women are less likely to become candidates 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. 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.

[16] For a broader survey, see Niederle and Vesterlund (2011).
[17] Fox and Lawless (2011).

Researchers attribute competition aversion to a variety of evolutionary and educational psychological explanations.[18] In all instances, though, "...the prospect of engaging in a future competition may cause women to anticipate a psychic cost and deter them from tournaments (Niederle and Vesterlund, 2007, p. 1070)" in a manner not experienced by men. In other words, women pay a higher fixed cost than men for choosing to engage in a competition, such as an election.

On its own, competition aversion cannot explain why female leaders may fight harder in conflicts. While it helps explain why men and women self-select into candidacy at different rates, it does not account for differences in how hard they fight in subsequent conflicts, *once they have been elected*. When making a decision, leaders have already overcome competition aversion by selecting into candidacy and winning election. Yet, they *still* behave differently. Women may feel competition aversion during a campaign or in the early stages of crisis bargaining, but once hostilities have escalated, competition aversion does not fully explain their choices.

A second theoretical component helps us explicitly link the decisions made before and after election. Individuals vary in their *non-monetary* value to winning contests. People of all genders vary with respect to how much psychological joy they feel from winning a contest, above and beyond any direct, monetary rewards.[19] This is distinct from competition aversion since it is only realized when the individual wins victory. We call this the participant's non-monetary value to winning (henceforth, "NMVW"). Having a high NMVW affects both the decision to enter an electoral contest and subsequent behavior in competitive situations. A higher NMVW helps overcome any competition aversion, which makes it more likely an individual will choose to enter an electoral contest. A higher NMVW also increases the marginal benefit of spending more on efforts to ensure victory in subsequent contests.

Competition aversion and heterogeneity in NMVWs *together* explain why female leaders may fight harder than their male counterparts, despite women being more pacifist in general. In line with research on competition aversion, we argue that women pay a larger fixed cost when entering a competition than men. To compensate for this, only women with a higher NMVW self-select

[18] Campbell (2013).
[19] Sheremeta (2010).

into leadership contests. This effectively screens out women with low NMVWs from leadership. Women with higher NMVWs also spend more on effort when they subsequently make choices in conflict situations on behalf of their electors.

Although men, like women, vary in their NMVW, the self-selection mechanism is weaker among men. Men generally face lower barriers to entry into elections since they do not have the same level of competition aversion. Women may be more attuned to the competitive nature of elections, creating a closer parallel between their preferences and behavior in elections and subsequent conflict situations. In line with this, Fox and Lawless (2011) find that women, in a survey of potential political candidates, are more likely than men to view elections as highly competitive. Accordingly, we expect the relationship between NMVW and selection into candidacy will be substantially stronger for women than men.

These gendered differences in NMVW could affect how candidates campaign for office. A woman with a high NMVW who decides to run is also likely to invest more effort into a successful campaign. In line with this, Anzia and Berry (2011) argue that "only the most talented, hardest working female candidates will win elections" and that "only the most qualified, politically ambitious females will emerge as candidates" (1). They then work harder on (co)sponsoring bills and delivering federal spending to their home districts. If female candidates with higher NMVW also put more effort into campaigning, and are therefore more likely to be elected, elections themselves 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. Women with higher NMVWs who opt into elections may exert greater effort on campaigning to cut through the noise.

We therefore expect that women with the highest NMVWs choose candidacy, work the hardest at the campaign stage and have a higher likelihood of winning the election. The gender gap in competition aversion makes this sorting mechanism significantly more influential for women than man. This leads to our main expectation:

Hypothesis. *Individuals with higher non-monetary values to winning contests are more likely to* *enter and win elections. This relationship is stronger for women than men.*

In short, women who fight hard in conflicts are particularly likely to become leaders because they hold preferences that also make them more likely to choose to run for office in the first place. This creates a pool of female candidates who are not representative of the broader citizenry when it comes to conflict preferences. To the best of our knowledge, the closest related observational studies to ours are Imamverdiyeva and Shea (2022) and Judge (2021). The former considers selection effects as a potential explanation for similarities in military spending between male and female leaders. They analyze lame duck female leaders and conclude that they cannot rule out the possibility that selection effects mean that male and female leaders have similar preferences over defense spending. Judge (2021) finds that many women rise to leadership after domestic crises, because they signal a change from the status quo and are viewed as capable of "righting the ship." Since domestic crises can persist and also trigger international crises, these conditions that create space for new female leaders may also be associated with conflicts during the subsequent leader's tenure.

3 Empirical Design

To assess our hypothesis, we face an important obstacle: we need to observe the decisions of participants that do *and do not* eventually become leaders. This is well documented in observational studies. For example, this issue motivates the comprehensive analysis of career backgrounds of female politicians in Baturo and Gray (2018). Bernhard, Shames and Teele (2020) track participants in a U.S. leadership training program to compare those that did and did not eventually run for office. This issue also motivates experimental work, where researchers can randomly assign features of a hypothetical leader's identity.[20] We overcome this obstacle by comparing the behavior of all participants in situations that mimic those faced by leaders, before and after a democratic election.

We recruited 162 participants for 10 sessions of our game from Amazon's Mechanical Turk

[20]Schwartz and Blair (2020).

(MTurk) in December 2019. Participants played in real-time. In each session, approximately 14 participants played one-on-one and intergroup lottery contest games. Lottery contest games model conflict in a tractable, understandable way.[21] As in war, players can decide to spend resources to win a zero-sum prize. In each contest, players start with an endowment of 1000 points that they can use to buy lottery contest tickets. Each player's likelihood of winning the contest equals the number of tickets bought divided by the total number of tickets bought by each player. The value of the prize varied each round, generally from 1225 to 2715 points. The winner of the contest received the prize and any unspent points from her endowment. The loser received only her unspent endowment points.[22]

At the start of each session, we randomly paired respondents in each of 12 rounds to play one-on-one lottery contest games. Participants were not identified and 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. We call this stage the Individual Contest Game (ICG).

Later, in what we call the Democratic Selection Game (DSG) stage, participants were randomly placed into either the "Blue Group" or the "Orange Group."[23] Group members remained anonymous, without any identifiers. We re-shuffled the groups every other round. In each round, groups chose their leaders via a democratic election. Participants first decided 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. Participants then voted for their group leader.[24] The winning candidate received a bonus of 245 points, which could not be used in the

[21] Chaudoin and Woon (2018).

[22] Appendix A gives a more detailed description of the protocol, sample, and compensation. We calculated a player's payment bonus at the end of each session, using their payoffs from 5 randomly selected rounds. Players did not accrue points over rounds, minimizing the effect of past winnings on current decisions. Participants knew this and were encouraged to think of every round as a separate decision task. Participants received a $5 show-up payment and $1 for every 210 points they won. Before playing, participants watched an animated video explaining the rules. We hired a graphics designer to make the instructional video clear and engaging.

[23] 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 possible dropouts, we sometimes added an extra player. In practice, dropouts were very rare. In Appendix B, we provide a detailed justification for this deception.

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

contest. This protocol reflects two key features of real-world elections: (1) participants must *self-select* into running and (2) there is some communication between candidates to make vote choice meaningful.

The two elected group leaders then played a contest game against one another, on behalf of their groups. The leaders chose how many tickets each group member would buy, which 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. In terms of expected utility, the leaders' decisions in the DSG were strategically identical to those made by individuals in the ICG. Any differences in behavior between the ICG and DSG are attributable to the presence of the election. Behavior in the ICG provides data about how each participant played individual contests, allowing us to see whether certain types of participants eventually select into candidacy and win elections, and whether this varies across gender.

To isolate the effect of self-selection, we purposefully ensured participant gender was unknown throughout. Individuals chose to be candidates, wrote campaign messages, and voted without knowing identifying characteristics of the other participants. To be clear, our claim is *not* that anonymity has erased all potential influences of gender or stereotyping. We acknowledge that even understanding that one's gender will not be explicitly revealed does not entirely remove all effects of gender bias. Internalized stereotypes could lead female leaders to act as though they have something to prove, even if they know their gender is unknown to others. Given the deep socialization of gender norms, there is no way to remove this entirely. However, using an anonymous laboratory setting allows us to get as close as possible to this ideal. Anonymity very clearly blocks some of the most common channels through which stereotyping generates differences between men and women leaders - e.g. through voter biases. Female participants were therefore "freer" to make choices without the overt influence of concerns about counter-stereotypical behavior or a fear of backlash against displaying masculine behavior. As shown below, we can also bring direct evidence to bear that selection effects are the main driver of our results, even if residual effects of stereotyping remained.

3.1 Measuring NMVW

We use two metrics of participant behavior to measure NMVW. First, we estimate each participant's NMVW based on how their behavior in the ICG compares to the Nash prediction of ticket purchases. 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 NMVW is therefore given by $v = 4T - p$.[25] For each participant, we calculate the average value of v, based on data from the ICG rounds, and call this the "Nash NMVW."

Second, we use the average number of tickets purchased in the ICG rounds. An individual with a higher NMVW will purchase more tickets than an individual with a lower NMVW. This measure thus reflects the quantity of interest without necessarily measuring it directly. It also has the advantage of not using Nash Equilibrium strategies as a benchmark. In practice, all the results are very similar for both measures.

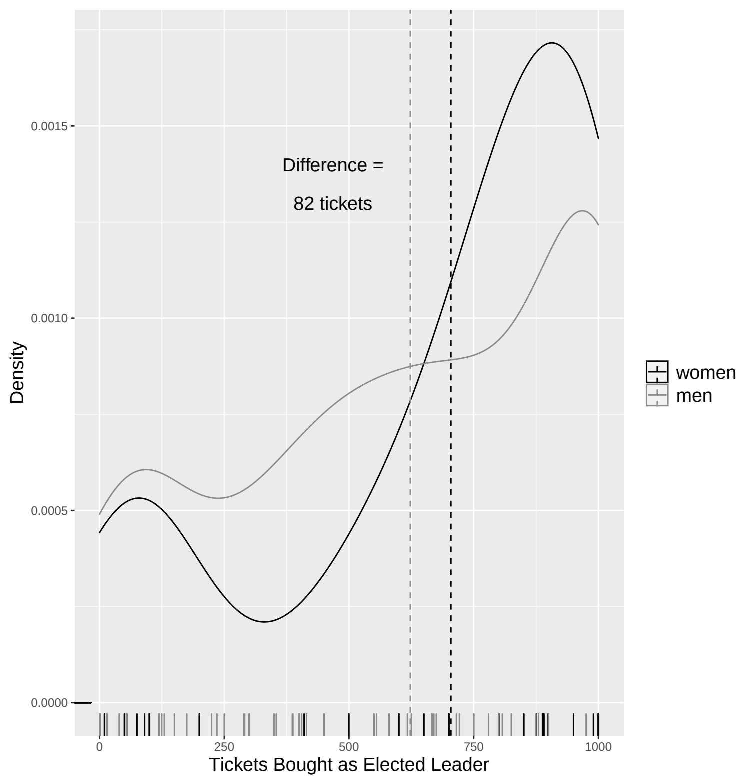
4 Results: The Power of Selection

4.1 Gender Differences in Leader Behavior

Our data reproduce the empirical phenomenon motivating many observational studies cited above: democratically elected female leaders spent substantially more on contests compared to male leaders, despite women spending less as individuals. Figure 1 shows the number of tickets bought by elected leaders in intergroup contests, split by gender. Female elected leaders purchased, on average, 82 more tickets (from a maximum possible 1000 tickets) per round than male leaders. This is a large increase in spending on conflict: it is over 8% of the total budget available. This difference

[25] Denote players i and j's ticket purchases as T_i and T_j. Player i's expected utility is $EU_i(T_i, T_j) = \frac{T_i}{T_i+T_j}(p + v) - T_i$. The players' first order conditions yields the Nash prediction.

Figure 1: Tickets Bought in the DSG, by Gender



arose despite the fact that women bought approximately 29 fewer tickets on average than men in one-on-one contests. These differences are statistically significant and robust to a wide array of regression specifications.[26]

4.2 Evidence of a Selection Effect

To demonstrate that selection explains these differences and provide evidence for our main hypothesis, we examine the difference in ICG behavior for leaders and non-leaders and show that this is much bigger for women than men. The top section of Table 1 compares the NMVW measures of elected leaders with those who were not elected, averaged across all DSG rounds and split by gender.[27] Among men, there is only a small difference in both measures between leaders and

[26]See Appendix C.

[27]A participant's ICG behavior "counts" toward the leader mean in rounds that s/he is a leader. This accounts for the fact that some participants became leaders more often by weighting the overall means accordingly.

non-leaders. Male leaders had a Nash NMVW of only 32 points higher and purchased just 8 more tickets per round, on average, in the ICG, compared to male non-leaders.

These differences are much larger for women. Female leaders had a Nash NMVW that was 598 points higher than female non-leaders. Female leaders purchased 150 more tickets on average in the ICG compared to female non-leaders. There is a stark difference in the degree of sorting among men and women, according to how much they value winning.

The largest part of the gap between NMVWs for female leaders versus non-leaders is explained by self-selection into candidacy, as shown by comparing the behavior of candidates and non-candidates in the middle portion of Table 1. Among men, there are very small differences between candidates and non-candidates. Male candidates had an average Nash NMVW that was just 14 points higher than that of male non-candidates, and they bought only 3 more tickets, on average, in the ICG, compared to male non-candidates (560 vs. 557).

Among women, however, these differences are again striking. Women who ran for elections had an average Nash NMVW that was 482 points higher than women who chose not to run. Female candidates purchased an average of 121 more tickets in the ICG, compared to female non-candidates (616 vs. 495). Self-selection effects are much larger for women than men.

Women with higher NMVWs were also more likely to win elections, conditional on candidacy, and this effect was stronger for women than men (bottom of Table 1). Election-winning men were virtually identical to losing male candidates. Winning male candidates had a 14-point higher Nash NMVW and purchased just additional 4 tickets in the ICG, compared to losing candidates. The differences for women were much larger. Winning female candidates had a Nash NMVW that was 317 points higher and purchased an average of 79 more tickets per round in the ICG. Since gender was not revealed at any stage, this difference in success is attributable to differences in campaigns, analyzed more below.

Each of the differences emphasized here is statistically significant at the 0.01 level. In the first two columns of Table 2, we regress (logit) an indicator for whether a participant was leader in a particular round of the DSG on the NMVW measures interacted with an indicator for women.

Table 1: Differences in NMVWs, by Gender

(I) Leaders vs. Not Leaders

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

(II) Candidates vs. Not Candidates

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

(III) Winning vs Losing Candidates

	Men		Women	
	winner	*loser*	*winner*	*loser*
Ave. Nash NMVW	522	508	979	662
difference	+14		+317	
Ave. ICG Tickets	561	557	675	596
difference	+4		+79	

Note: Part III excludes candidates who ran unopposed.

Table 2: Effect of NMVW on Leadership and Candidacy, by Gender

	Dependent variable:					
	Is Leader		Is Candidate		Is Winner	
	(1)	(2)	(3)	(4)	(5)	(6)
					(candidates only)	
Female	1.427	3.166	0.958	2.473	0.940	1.991
	(0.237)	(0.570)	(0.130)	(0.336)	(0.288)	(0.686)
Nash NMVW	0.005		0.002		0.002	
	(0.010)		(0.008)		(0.012)	
Female Nash NMVW	0.101		0.088		0.061	
	(0.022)		(0.014)		(0.026)	
Ave. ICG Tickets		0.019		0.008		0.008
		(0.041)		(0.031)		(0.049)
Female Ave ICG Tick.		0.404		0.352		0.244
		(0.087)		(0.056)		(0.104)
Constant	1.625	1.708	0.286	0.321	0.706	0.739
	(0.099)	(0.244)	(0.074)	(0.183)	(0.121)	(0.291)
Observations	1,786	1,786	1,786	1,786	645	645

Note: p<0.1; p<0.05; p<0.01
Models (5) and (6) exclude unopposed candidates.

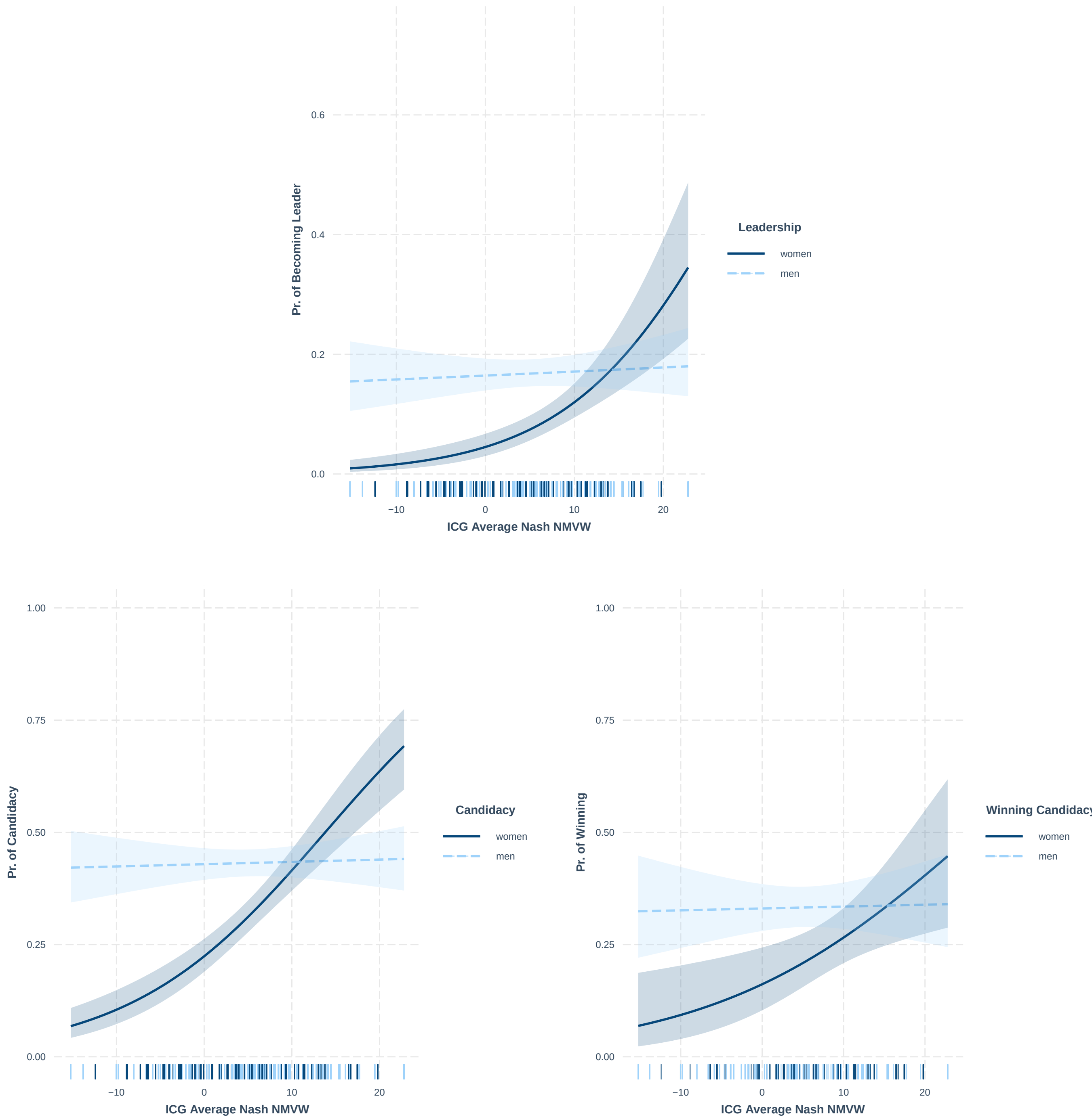
In the second two columns, we do the same using an indicator for whether a participant self-selected into candidacy. In the final two columns, we do the same using an indicator for whether a particular candidate won election. In each, the interaction term is substantively and statistically significant, showing that NMVW has a stronger positive effect on the probability of being a leader or a candidate for women than men.[28]

Figure 2 shows the estimated effects visually. The top panel illustrates how NMVW affects the probability of becoming leader for men and women. The bottom left panel shows the same for the probability of being a candidate; the bottom right shows the same for the probability of being a winning candidate.

The results are consistent and striking. Only women who are among the highest in their NMVWs run for and win office at rates comparable to men. They then proceed to fight harder in inter-group competitions. Even in a simplified laboratory setting - in which gender markers are absent - the selection mechanism highlighted in this article is strong enough for the emergence of an overall pattern wherein female leaders fight harder.

[28]We re-scaled the NMVW measures to be in 100's of points. All results are similar using OLS and in a wide array of robustness checks; see Appendix D.

Figure 2: Predicted Effects of Nash NMVW, by Gender



Note: The plot shows predicted probability of being a leader or candidate in a particular round of the DSG. Predictions use estimates from Table 2, columns 1 and 3.

4.3 Gender and Campaign Strategies

The previous section provided evidence for the main result - that selection dynamics are much more stark for women than men. Here, we turn to a deeper exploration of how high-NMVW women succeeded in getting elected at a higher rate. We posited that women with higher NMVWs will work harder at the campaign stage and subsequently are more likely to be elected. Unlike the decision to run, which is a simple yes-or-no decision, the choice to work harder in the campaign stage is more complicated to observe and measure. Here, we demonstrate that women with higher NMVWs wrote longer campaign messages - a reasonable proxy for effort - which helped them win elections. Additionally, even accounting for message length and content, higher NMVW women wrote more effective campaign messages. These relationships were less strong for men.

The effort put into writing a message is difficult to measure, but in our protocol we do have one measure that has construct validity: message length. Writing longer messages takes additional time and cognitive energy.[29] "Turkers" — people doing tasks on Mturk — prioritize task speed and want to maximize their wages per hour. Websites where Turkers compare and review potential tasks highlight the effective pay per hour. Browser extensions help Turkers scrape for higher-paying, shorter tasks, with links to reviews showing the task's pay per hour.[30]

Women with higher NMVWs wrote longer campaign messages, which in turn, increased their success rates. Table 3 shows results from regressing message length on our NMVW measures, an indicator for female candidates, and their interaction. Women with higher NMVWs wrote substantially longer messages, which was not the case for men. Longer messages also were more successful in campaigns. An additional 50 characters in a message raised a candidate's probability of winning by approximately 13%.

Formal mediation analysis provides additional evidence that higher NMVW women wrote longer messages which helped them win. We used message length as the mediator, our NMVW measure as the treatment, and whether the candidate won as the outcome of interest. For women,

[29] Giovannoni and Feltovich (2022).

[30] Eg `turkopticon.net`. We do not have a measure of the time they spent entering their message. We don't have any reason to think results would differ, though.

the average causal mediation effect was positive and significant (ACME estimate = 0.02, p= 0.01).[31] The message length mediator accounted for approximately 15% of the total effect of NMVWs on campaign success for women. For men, this was not the case. Higher NMVW men actually wrote slightly shorter messages, decreasing their electoral success.

Table 3: 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)	
Nash NMV		0.586
		(0.219)
Female Nash 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.

While length is a reasonable proxy for effort, it lacks nuance. We would ideally like a measure

[31] Since NMVW is continuous, this estimate corresponds to moving from the 25th to the 75th percentile of NMVW. Estimates are from the method described in Tingley et al. (2014).

of message *quality* or *persuasiveness*, which likely increase with the effort put into crafting a message. Measuring message quality is obviously subjective and difficult. It is further complicated by the fact that men and women, and participants with different NMVWs, can (and do) chose different content in their electoral appeals. Persuasiveness can vary a lot across different message types and messages with similar content can still vary widely in their persuasiveness.

Here, we show evidence that higher-NMVW women likely crafted higher quality messages — even when accounting for message length and message content. This pattern does not appear for men.

We coded each message as being one of ten message types. The message types are given below, with an actual example for each.[32] We order the categories from those that were most to least successful.

Campaign Message Types

- Strategy: a statement about the strategy the candidate will use as leader. Eg "I will bet approximately 100 tickets for each 500 in the prize."
- Track record: an appeal to the candidate's past success. Eg "I won the last three rounds where I was leader and can do it again."
- Speed: a promise to make quick decisions. Eg "I will make fast decisions to get this game moving faster."
- Bid high: an appeal for higher or riskier bidding. Eg "I like higher bids go big or go home."
- Critique: a negative appeal against past leaders' decisions. Eg "Don't pick that dude again. I can win for us."
- Skill: an appeal to the candidate's skill as a leader. Eg "I understand the game well and will get us the most money."
- Team: an appeal to team identity or against the out-group. Eg "let's take down that pesky orange team!"

[32] See Appendix E for summary data about message types.

- Bid low: an appeal for lower or more conservative bidding. Eg "I wont go over 200 we will get 800 at the least."
- Humor: the candidate used humor. Eg "baby yoda for president."
- Null: a small subset of messages that didn't fit into the above categories.

Table 4 shows the effect of NMVW on the probability of winning election, for men and women, controlling for message length (column 1), message type (column 2), and both (column 3). In all specifications, a higher NMVW increases the probability of winning for women, as evidenced by the positive and significant interaction terms. Higher NMVW women are not simply writing certain types of message; rather, controlling for the broad message category, they are writing *better* messages. Likewise, they are not only writing longer messages than low NMVW women; they are also writing better ones. These findings are much weaker for men.[33]

[33]Messages can have content that fits with multiple categories. Here, we coded the message's primary category. Results are similar using coding for whether the message has any content that fits a particular category. See Appendix E.

Table 4: Effect of Gender and NMVW on Winning, Controlling for Campaign Messages

	Dependent variable:		
	Wins Election		
	(1)	(2)	(3)
NMVW	0.008	0.003	0.007
	(0.013)	(0.014)	(0.014)
Female	0.825	0.925	0.861
	(0.300)	(0.319)	(0.329)
NMVW Female	0.053	0.069	0.062
	(0.027)	(0.030)	(0.030)
Message Length	0.012		0.009
	(0.002)		(0.002)
Message Content:			
Bid High		1.350	1.035
		(0.714)	(0.725)
Bid Low		0.854	0.531
		(0.784)	(0.792)
Humor		0.540	0.359
		(0.664)	(0.669)
Critique		1.447	1.136
		(0.677)	(0.684)
Skill		1.437	1.093
		(0.633)	(0.642)
Speed		2.156	1.973
		(0.848)	(0.850)
Strategy		2.310	1.837
		(0.722)	(0.736)
Team		1.054	0.837
		(0.674)	(0.680)
Track Record		2.259	1.881
		(0.645)	(0.654)
Constant	1.533	2.110	2.441
	(0.211)	(0.624)	(0.632)
Observations	638	638	638
Log Likelihood	371.953	361.449	354.799
Note:	p<0.1;	p<0.05;	p<0.01

5 Alternate Explanations

5.1 Does risk aversion explain our results?

One alternative explanation is that risk preference, which potentially differs across gender, explains self selection into candidacy. For example, it is theoretically possible that only the most risk tolerant women select into candidacy and then purchase additional tickets. We can use the data from the ICG to demonstrate that this does not explain our results.

It is first necessary to understand how risk aversion affects ticket purchases as the prize value changes. As the prize value increases, every participant, regardless of their risk preferences, should buy more tickets. However, the marginal increase in ticket purchases should be higher among the more risk tolerant participants.[34] For a risk neutral player, under the Nash prediction, a 100 point increase in the prize value should cause the participant to buy 25 more tickets. For a risk averse participant, their ticket purchase should increase buy fewer than 25 tickets. For a risk loving participant, the increase in the prize should cause her to increase her ticket purchase by more than 25 tickets. The intuition is straightforward: the risk acceptant participant is enticed to gamble as the prize increases more so than the risk averse participant.

In the ICG, we included two low value rounds where the prize was only worth 275 or 280 points. The next highest prize values are 1225 and 1235. For each participant, we calculated their average ticket purchase when the prize was 275 or 280 points and their average ticket purchase when the prize was 1225 or 1235. We then used calculated their average percent increase in ticket purchases as the respondents moved from the low values to the high values. This gives a measure of risk preferences across participants: the percent increase should be higher among more risk acceptant participants.

Table 5 replicates Table 1 using this measure. The results do *not* suggest that more risk acceptant women become leaders, select into candidacy, or win elections.

Among women, leaders' ticket purchases in the 1225/1235 point rounds was approximately

[34] Note that the relationship between tickets purchased and risk preferences is complex. Empirically, more risk averse participants in contest games purchase fewer tickets.

7.65 times larger than their purchase in the low value rounds (top section). Women who did not become leaders increased their purchases to a much greater degree. Female non-leaders actually appear more risk acceptant than female leaders. The same is true when looking at candidates versus non-candidates (middle section). Women candidates increased their ticket purchases to a smaller degree than those who were not candidates. The same is again true for winning versus losing female candidates, where the winning candidates tended to be women who increased their ticket purchases to a lower degree (bottom section).

Table 5: Differences in ICG behavior, Risk Aversion (LVR and 1225 rounds), by Gender

(I) Leaders vs. Not Leaders				
	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Av. Perc change in tickets	35.75	20.49	7.65	17.64
(II) Candidates vs. Not Candidates				
	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Av. Perc change in tickets	28.91	19.51	10.43	16.96
(III) Winning vs. Losing Candidates				
	Men		Women	
	winner	*loser*	*winner*	*loser*
Av. Perc change in tickets	28.46	24.65	7.27	11.67

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

We chose to use the 1225/1235 values because they are the next highest values. They are also prize values that are only marginally above the ticket purchase cap of 1000 tickets. At higher prize values, the ticket cap is more likely to bind, making it harder to assess candidates' risk preferences. However, our results - and their inconsistency with a story based on risk aversion - obtain even when we compare other pairs of prize values.[35] The appendix also shows how our main results -

[35] See appendix.

about the differential effect of NMVW for women - obtain, even when controlling for these risk aversion measures.

5.2 Does confidence explain our results?

As with risk aversion, it is possible that there is self-selection based on participants' confidence, which also potentially differs across genders. The underlying task in our game is *not* one where participants have a directly measurable skill level. It is strategic; a participant's payoff depends on her choice and that of her opponent. Payoffs also have a stochastic element. Therefore, the best information a participant receives that can shape her confidence level is her actual payoffs. If a participant receives more points, 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 the outcomes of the contest, ticket purchases, and payoffs, after every round.

For each participant, we calculate their average payoff in the ICG. All else equal, winning more points in the ICG should increase a participant's confidence in her strategic decision-making. This is meaningful data that each participant gets about the fruitfulness of their strategies.

Leaders in the DSG tended to have higher ICG payoffs than non-leaders.[36] 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 and winning versus losing candidates. This is evidence that selection was not driven by confidence. In the appendix, we also show analysis matching the analysis of risk aversion above, giving further evidence that confidence in contests and confidence in one's ability to win an election does not explain our results.

[36]Table omitted for length; see appendix.

6 Conclusion

To become a chief executive, one must choose to run and then win a competitive election. The electoral gauntlet is not for the faint of heart. As a result, the very things that determine whether women choose to run and succeed in an election also mean female leaders are more willing to spend resources in order to win contests. Consistent with this, we demonstrate that female elected leaders invest more in costly competition than their male counterparts in a controlled laboratory setting. Women who choose to run and then eventually win elections are among those who most highly value winning for winning's sake. Such selection effects are not nearly as stark for men. We found these gendered patterns despite minimizing the effects of gender-bias and stereotyping by making participants anonymous.

The implication of this is that we cannot attribute the decisions of female leaders solely to the strategic constraints placed on them by their environment. While sexism and stereotypes shape the choices of female leaders, selection mechanisms mean that those choices also reflect the leader's held and/or learned preferences and beliefs. And these selection mechanisms operate beyond the unequal electoral constraints placed on women by voters who expect candidates to conform to stereotypes about their gender. Self-selection into candidacy and choices over how to campaign are especially important.

A deeper dive into specific leaders and would-be leaders is beyond the scope of this note. But in-depth analysis of some of the female leaders most associated with international conflict shares our arguments. Writing about Meir, Gandhi, and Thatcher, Blema Steinberg (2008) argues that "As important as the singular effects of gender are on leadership behavior, they should not be overstated. ... the numerous aspects/patterns of the personalities developed through [the three leaders'] earlier life experiences were carried into the prime ministerial office and translated into their particular leadership styles" (11). Biographies of these leaders are replete with anecdotes of their intense drive to win and succeed. Related research into foreign policy leaders emphasizes how their " patterns of assertive behavior ... pre-date[d] recruitment to senior posts, suggesting that the repertoires of foreign policy leaders [were] in place before they [reached] executive office."[37]

While selection mechanisms have been speculated to play a role in the choices of female leaders, our research helps establish micro-foundations for these arguments. We develop a theoretical framework that helps better specify what traits are being selected on both by potential candidates and electorates, and how these traits then affect leader decisions. Beyond just saying that female leaders are non-representative of the general population, our research specifies the exact dimensions that are driving non-random selection and leader decisions. If the external constraints of stereotyping and sexism become less binding for female leaders over time (or so we hope), then the explanatory power of characteristics like a leader's intrinsic drive to win contests may go further and further towards explaining variation across leaders.

Future research could compare and contrast offices with different selection mechanisms. Barnes and O'Brien (2018) show how a state's internal and external security situation affect the likelihood of women being appointed to defense ministries. Those state characteristics also affect the likelihood of future interstate conflict, which confounds estimates of the effect of minister gender on conflict. There could also be different self-selection patterns by gender if selection by appointment differs from an electoral gauntlet. The appointment process may be (partially) shielded from public scrutiny and biases of the electorate. But it is also a competitive gauntlet as well, as candidates jockey behind the scenes for the favor of those in control of appointments or for a more powerful cabinet position. Self-selection could potentially play a large role in these domains, as well.

[37]Bashevkin (2018) p 2.

References

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.

Bashevkin, Sylvia. 2018. *Women as foreign policy leaders: national security and gender politics in superpower America*. Oxford University Press.

Baturo, Alexander and Julia Gray. 2018. "When do family ties matter? The duration of female suffrage and women's path to high political office." *Political Research Quarterly* 71(3):695–709.

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.

Blair, Christopher W and Joshua A Schwartz. 2021. "The Gendered Peace Premium." *Working paper* .

Burns, Courtney and Jeremy Bowling. 2021. "Signaling Woman and Leader: Navigating the Double Bind as a Foreign Policy Decisionmaker." *Journal of Women, Politics & Policy* 42(4):332–351.

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

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.

Cohen, Dara Kay and Sabrina M Karim. 2021. "Does More Equality for Women Mean Less War? Rethinking Sex and Gender Inequality and Political Violence." *International Organization* pp. 1–31.

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

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.

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.

Giovannoni, Francesco and Nicholas Feltovich. 2022. "Campaign messages, polling, and elections: theory and experimental evidence." *American Journal of Political Science* .

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

Imamverdiyeva, Ulkar and Patrick E Shea. 2022. "Re-examining women leaders and military spending." *Journal of Peace Research* p. 00223433211055909.

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.

Kertzer, Joshua D. Forthcoming. "Re-Assessing Elite-Public Gaps in Political Behavior." *American Journal of Political Science* .

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't a 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).

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

Reiter, Dan and Scott Wolford. 2022. "Gender, sexism, and war." *Journal of Theoretical Politics* 34(1):59–77.

Renshon, Jonathan, Julia J Lee and Dustin Tingley. 2017. "Emotions and the micro-foundations of commitment problems." *International Organization* 71(S1):S189–S218.

Sahm, Marco. 2017. "Risk aversion and prudence in contests.".

Schramm, Madison and Alexandra Stark. 2020*a*. "Peacemakers or Iron Ladies? A Cross-National Study of Gender and International Conflict." *Security Studies* 29(3):515–548.

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

Schwartz, Joshua A and Christopher W Blair. 2020. "Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining." *International Organization* pp. 1–24.

Sheremeta, Roman M. 2010. "Experimental comparison of multi-stage and one-stage contests." *Games and Economic Behavior* 68(2):731–747.

Steinberg, Blema S. 2008. *Women in power: The personalities and leadership styles of Indira Gandhi, Golda Meir, and Margaret Thatcher*. Vol. 4 McGill-Queen's Press-MQUP.

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.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai. 2014. "Mediation: R package for causal mediation analysis.".

Trager, Robert F and Joslyn N Barnhart. 2023. *The Suffragist Peace: How Women Shape the Politics of War*. Oxford University Press.

Appendices for "Elections, War, and Gender: Choose to Run, Choose to Fight"

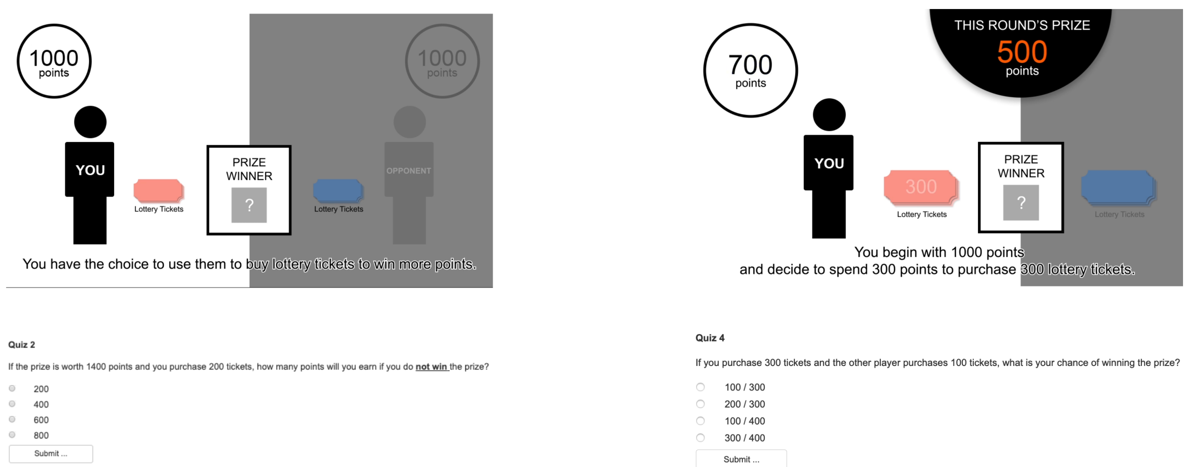
A Appendix: Full Protocol Description, Recruitment, Demographics

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

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.[38] 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



We also included a part of the game after the ICG but before the DSG where participants were randomly sorted into groups and a group leader was randomly chosen. We called this the "Random Selection Game" (RSG). This part was also 12 rounds long with a similar prize value sequence, shown below. (Note: all of our results about selection based on calculations from behavior in the ICG also obtain using data from the RSG. These are omitted from the appendix because of journal page limits, but are available on request.)

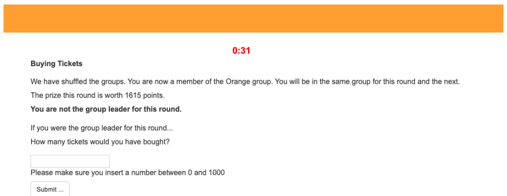
The top two panels of Figure A.2 show screen shots from the game. The top left panel shows an example of a player who was not chosen as group leader and was then asked how many tickets she would have purchased. The screen for a group leader looks similar. The top right panel shows an example of what 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

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

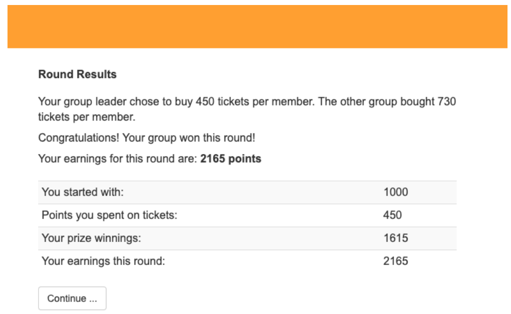
earnings for that round.[39]

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

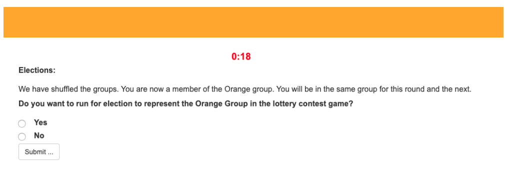
Figure A.2: Screen Captures from the Random Selection and Democratic Selection Parts of the Game



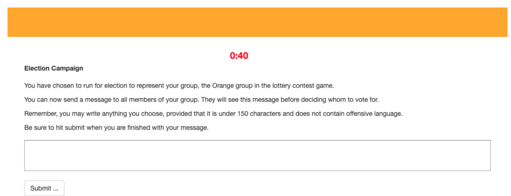
(a) Ticket Purchase



(b) Round Results



(c) Candidate Decision



(d) Campaign Message

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.[40]

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

[40]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 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 Compensation

All participants received a $5 show up payment. If a session was full and they didn't play the game, they still received this payment. Bonus amounts were calculated at an exchange rate of 210 points per $1. We randomly chose 5 rounds, excluding ones where every participant was told they were a group leader, calculated the average winnings from that round, and used that for the bonus amount.

A.3 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 filled out a brief pre-survey that was posted an hour before the start time. They had to pass a reCAPTCHA screen and agree to show up a pre-designated time before submitting the pre-survey. We messaged an individualized study link and instructions to those who had completed the survey. If the participant didn't show up at the designated time, they couldn't participate.[41] For those who did show up

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

on time and participated, 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 and paid a prorated bonus for the time spent on the study. Participants were placed in the virtual waiting room (for up to about a minute) after each round while waiting for others to finish making decisions. As they don't know when they would be released from the waiting room for the next round, they had to be attentive in order to successfully complete the game without being excused.

A.4 Demographics of MTurk Sample

Our sample was closer to the U.S. 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 they reported a larger number of demographic characteristics than most studies using student samples. Compared to the CCES, our MTurk participant pool was about 10.5 years younger while the university laboratory pool was about 27.2 years younger. 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.[42] 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 sample had 12.5 percent more.

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

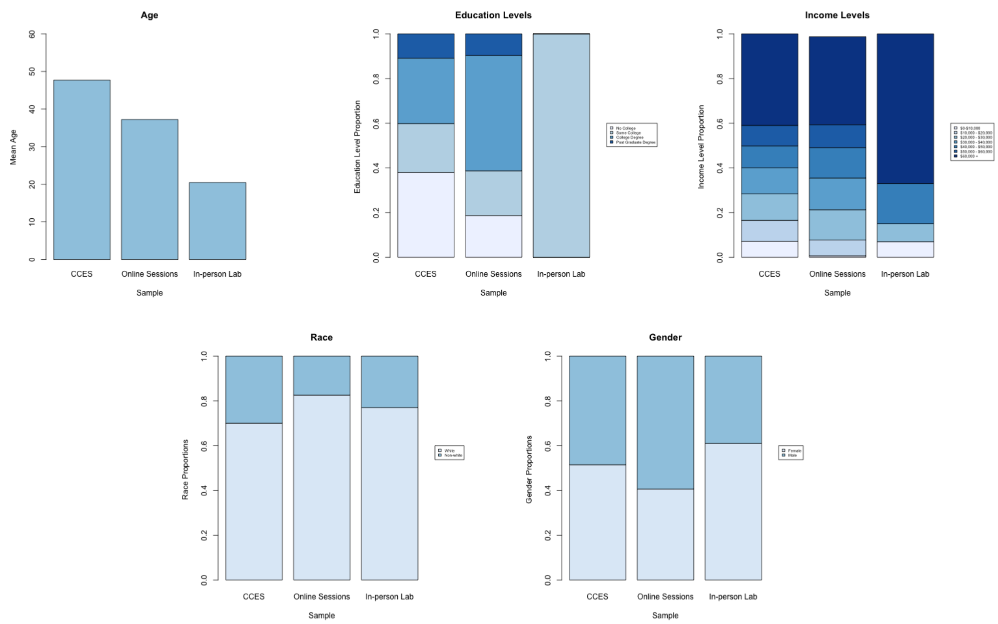


Figure A.3: Comparison of Sample Characteristics

B Appendix: Deception Justification

Our 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-session 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 an alternate protocol would not have been feasible.

B.1 Pre-brief and Debrief

We made respondents aware of the possibility of deception before consenting to participate. Pre-brief about the possibility of deception 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):

> ***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.

B.2 Economic or Material Costs; Cognitive or Psychological Trauma

If any respondent felt uncomfortable with the possibility of deception at the informed consent stage, they could decline to participate in the research 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.

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.

B.3 Full Randomization

An 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

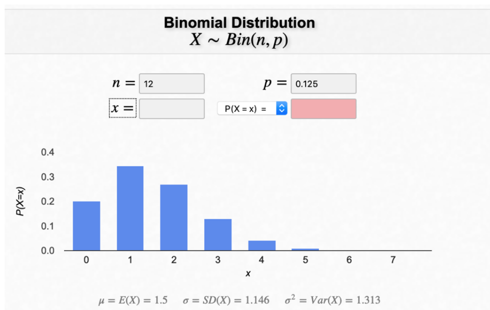


Figure Source: Bognar, Matt. Department of Statistics and Actuarial Science, University of Iowa. `https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html`.

The subject payments for these sessions 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: Differences in Leader Behavior by Gender

In Table C.1 we show results from regressing the number of tickets bought on indicators for whether the round took place in the DSG (as opposed to the ICG) and whether the participant was female and the interaction between the two. The regression coefficients therefore show all of the information implied by our comparisons. The Female constituent terms are all negative and significant: women bought fewer tickets in the ICG. The interaction terms are all positive and significant: women bought more tickets in the DSG.

The first column replicates the comparison and p-value noted in Figure 1. The second column shows how the results are even stronger when excluding the low-value rounds (LVRs), where the prize was less than 1,000 points. The final three columns exclude one female participant who was a frequent leader, the male participants who were most frequently leaders, and then excluding both of those groups.

Since men and women were leaders in different round and those rounds may have had different prize values, replicates that same series of regressions and includes prize value fixed effects. The estimated increase in tickets bought by female leaders is generally bigger in these specifications.

Table C.1: Effect of Gender on Tickets Bought

	Dependent variable:				
	Incl. LVR	No LVR	Excl. Fem. Outl.	Excl. Male Outl.	Excl. M/F Outl.
	(1)	(2)	(3)	(4)	(5)
Female	28.923	39.918	35.463	27.646	34.186
	(16.310)	(15.309)	(16.343)	(16.470)	(16.500)
DSG	61.314	60.346	61.314	57.133	57.133
	(27.585)	(25.779)	(27.506)	(30.176)	(30.085)
Female X DSG	110.420	155.654	96.073	114.600	100.253
	(52.466)	(49.666)	(54.474)	(53.883)	(55.827)
Constant	561.522	652.017	561.522	560.246	560.246
	(10.401)	(9.762)	(10.371)	(10.646)	(10.614)
Observations	2,040	1,700	2,021	1,960	1,941
Note:					p<0.1; p<0.05; p<0.01

Table C.2: Effect of Gender on Tickets Bought, with Prize Fixed Effects

	No LVR	Incl. LVR	Excl. Fem. Outl.	Excl. Male Outl.	Excl. M/F Outl.
	(1)	(2)	(3)	(4)	(5)
Female	39.918	28.923	35.463	27.646	34.186
	(14.652)	(12.813)	(12.813)	(13.057)	(13.056)
DSG	60.762	57.221	57.228	50.503	50.522
	(24.677)	(21.673)	(21.567)	(23.932)	(23.815)
Female X DSG	154.021	126.011	112.302	132.404	118.703
	(47.564)	(41.238)	(42.730)	(42.752)	(44.209)
Constant	523.295	124.143	124.963	128.091	128.992
	(16.437)	(15.542)	(15.516)	(16.028)	(16.004)
Observations	1,700	2,040	2,021	1,960	1,941
Note:					$p<0.1$; $p<0.05$; $p<0.01$

D Appendix: Robustness of Effect of NMVW by Gender

D.1 Leadership and Candidacy

Table 2 in the main manuscript showed the effect of our NMVW measures on whether a participant became a leader and whether they became a candidate, broken down by gender. The table showed the NMVW had a much larger effect on both for women than for men. Here, we show a variety of robustness checks for those results, again using the NMVW measures scaled in 100s of points.

We first replicated the logit regressions using OLS. We then replicated results using the NMVW measures that excluded LVRs from their calculations. We then replicated results excluding one female participant who was a frequent leader, frequent male leaders, and then both groups. Table D.1 shows results from these regressions using only the Nash NMVW measure, using an indicator for whether the participant was a leader (left side) and a candidate (right side).

As above, all of the interaction terms are positive and significant, showing the our NMVW measures increase the likelihood of becoming leader and being a candidate, more so for women than men. The only things presented in the main manuscript that were affected in meaningful ways by these specification decisions were those showing the relationship between NMVW and the likelihood of winning, conditional on being a candidate. There, excluding the one female participant decreases the statistical significance of the interaction term, though it is still positively signed.

Table D.1: Effect of NMVW on Leadership and Candidacy, by Gender

	Is Leader				Is Candidate			
	OLS	No LVR	No F Outl.	No M/F Outl.	OLS	No LVR	No F Outl.	No M/F Outl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.192	0.950	0.938	0.850	0.111	1.411	1.357	1.191
	(0.026)	(0.130)	(0.129)	(0.130)	(0.019)	(0.240)	(0.232)	(0.235)
Nash NMVW	0.001		0.002	0.002	0.001		0.005	0.002
	(0.002)		(0.008)	(0.008)	(0.001)		(0.010)	(0.011)
Fem. X NMVW	0.018		0.079	0.079	0.007		0.083	0.086
	(0.003)		(0.014)	(0.014)	(0.002)		(0.023)	(0.023)
NMVW (no LVR)		0.001				0.007		
		(0.007)				(0.009)		
F X NMVW (no LVR)		0.079				0.089		
		(0.012)				(0.020)		
Constant	0.429	0.283	0.286	0.374	0.165	1.642	1.625	1.791
	(0.017)	(0.074)	(0.074)	(0.075)	(0.012)	(0.100)	(0.099)	(0.106)
Observations	1,786	1,786	1,774	1,726	1,786	1,786	1,774	1,726

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

D.2 Winning Candidate

Table 1 in the main manuscript showed how higher NMVWs were associated with a greater likelihood of winning, conditional on candidacy, and that this effect was much stronger for women. The table just showed a comparison of the NMVW measures by winning and losing candidates, by gender.

Those comparisons are statistically significant. We used an indicator for whether a candidate won election as the outcome variable. We regressed that on an indicator for female candidates, the NMVW measures, and their interactions. We excluded observations where the candidate ran unopposed. The positive and significant interaction terms indicated that a higher NMVW was associated with a greater probability of winning and that this effect was larger for women than men.

We then replicated those same regressions only excluding LVRs from the calculation of the NMVW measures and excluding the outlier participants who were frequent leaders, respectively. The results are similar, though less strong statistically. All results again omitted.

E Appendix: Campaign Messages Coding and Analysis

Table E.1 shows the frequency and success rates of each type of message. Each column corresponds to a particular message type. We ordered them from least successful to most successful, with the percentage of times that message type won listed in the top row. For each message type, we then listed — by women/men — the number of times that message was used, the percentage of times that message type was used, the number of individual participants that used that message at least once, and the average NMVW of the participants using that message.

Women and men tended to campaign differently and the relationship between NMVW and message type also varied across gender. Women were less likely to use humor in their messages and more likely to appeal to a sense of team than men. Higher NMVW women were more likely to choose messages that appealed to their track records and appealed to a sense of team. Higher NMVW men were more likely to appeal to their own skills. High NMVW participants of both genders were, unsurprisingly, more likely to write messages promising to make higher bids.

Table E.1: Differences in Campaign Messages, by Gender and NMVW

	Null	Humor	Bid Low	Team	Skill	Critique	Bid High	Speed	Record	Strategy
Win Percent	9%	17%	18%	25%	31%	31%	32%	38%	52%	55%
Women										
Primary Cat.	17	24	10	46	44	22	24	9	24	0
	8%	11%	5%	21%	20%	10%	11%	4%	11%	0%
Individuals	11	12	5	17	24	13	8	3	13	0
Ave. NMVW	402	942	-156	1024	360	794	1269	210	876	NA
Men										
Primary Cat.	17	83	23	27	118	32	13	4	72	29
	4%	20%	6%	6%	28%	8%	3%	1%	17%	7%
Individuals	10	37	9	21	45	23	10	3	30	9
Ave. NMVW	908	435	-767	644	836	605	932	769	339	271

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).

The results also obtain if we use a coding of message types based on whether the message contained any amount of a particular type, as opposed to being the message's primary type. Table E.2 replicates Table 4 from the main manuscript, using this coding. We again find that, for women, higher NMVWs are associated with an increased chance of winning, compared to men, even when controlling for this alternate message type coding.

Table E.2: Effect of Gender and NMVW on Winning, Controlling for Campaign Messages (Any content coding)

	(1)	(2)
NMVW	0.005	0.008
	(0.014)	(0.014)
Female	0.926	0.871
	(0.318)	(0.326)
Message Length		0.008
		(0.003)
Bid High	0.522	0.540
	(0.471)	(0.468)
Bid Low	0.959	0.941
	(0.509)	(0.504)
Humor	0.013	0.126
	(0.301)	(0.307)
Critique	0.707	0.533
	(0.324)	(0.331)
Skill	0.558	0.393
	(0.262)	(0.272)
Speed	0.671	0.614
	(0.509)	(0.514)
Strategy	1.252	0.998
	(0.392)	(0.400)
Team	0.268	0.076
	(0.200)	(0.211)
Track Record	0.963	0.852
	(0.249)	(0.254)
NMVW x Female	0.062	0.060
	(0.029)	(0.030)
Constant	1.525	1.824
	(0.281)	(0.302)
Observations	638	638
Log Likelihood	361.796	356.977

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

F Appendix: Alternative Explanations

In this section of the appendix, we consider several alternative explanations for our findings. First, we show how our results are more consistent with a theoretical explanation based on non-monetary values to winning, as opposed to gendered differences in risk aversion. We do this by leveraging comparisons in ticket purchases across different prize values. It is not the case that results are explained by the most risk-acceptant women running, winning, and then buying more tickets as leaders.

Second, the DSG portion of the game adds an election, which potentially induces two changes that are distinct from selection effects. The experience of being elected can affect behavior directly (Park, Hummel and Chaudoin, 2022). This "election effect" could vary across gender. We consider this first and do not find any evidence that the election effect varies by gender.

Additionally, 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.

Finally, we show that our results are not consistent with gendered differences in confidence.

F.1 Risk Aversion

As noted in the main manuscript, the relationship between risk aversion is complex. Greater risk aversion might make a respondent purchase fewer tickets to keep more of her endowment *or* it could make her purchase more tickets to better guarantee a win.[43] Empirically, the first effect tends to dominate, so we focus on that effect here.

The main manuscript described why we chose particular prize values to assess risk aversion across participants. Here, we can show these same results, when calculating the percentage increase in tickets using the LVRs and the rounds with 1605/1615 prize values. We can also show this using the percentage change moving from rounds with 1225/1235 values to 1605/1615 values. In all cases, women leaders, candidates, and winning candidates showed lower percentage increases in their ticket purchases, which is inconsistent with selection on risk tolerance.

Table F.1 shows the same table as above, only we calculated the percentage increase in tickets using the low value rounds and the rounds with 1605/1615 prize values. Table F.2 again shows the same analysis, only using the percentage change moving from rounds with 1225/1235 values to 1605/1615 values.

The percentage increases are also generally what we would expect. Keep in mind that the jump from 275 points to 1225 or 1605 is a large increase, so we would expect large increases in ticket purchases, in percentage terms. The jump from 1225 to 1605 is smaller, and the percentage increase is correspondingly smaller in Table F.2.

Table F.3 shows statistical analysis demonstrating that risk aversion does not explain our results. We regressed an indicator for whether the participant chose to be a candidate on the risk aversion measures, our NMVW measure, and their interactions with a female indicator. In all cases, we still find that higher-NMVW women were more likely to run (bottom interaction term), even when

[43] Sahm (2017).

Table F.1: Differences in ICG behavior, Risk Aversion (LVR and 1605/1615 rounds), by Gender

(I) Leaders vs. Not Leaders				
	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Av. Perc change in tickets	51.41	37.26	7.65	17.67
(II) Candidates vs. Not Candidates				
	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Av. Perc change in tickets	46.96	35.53	11.99	19.95
(III) Winning vs. Losing Candidates				
	Men		Women	
	winner	*loser*	*winner*	*loser*
Av. Perc change in tickets	41.02	44.19	7.80	13.58

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

Table F.2: Differences in ICG behavior, Risk Aversion (1605 and 1225 rounds), by Gender

(I) Leaders vs. Not Leaders				
	Men		Women	
	leader	*not leader*	*leader*	*not leader*
Av. Perc change in tickets	0.25	0.22	0.05	0.11
(II) Candidates vs. Not Candidates				
	Men		Women	
	candidate	*not candidate*	*candidate*	*not candidate*
Av. Perc change in tickets	0.22	0.23	0.06	0.13
(III) Winning vs. Losing Candidates				
	Men		Women	
	winner	*loser*	*winner*	*loser*
Av. Perc change in tickets	0.24	0.20	0.04	0.06

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

controlling for risk aversion measures. Results are similar when replicating these regressions using an indicator for whether a candidate won as the dependent variable. Higher NMVW women candidates are more likely to win, even controlling for risk aversion measures. (Results omitted for length and are available on request).

F.2 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. We regressed the number of tickets purchased on indicators for the DSG part of the game. We included participant fixed effects, so that we are estimating within-participant changes in ticket purchases across different parts of the game. As above, we included prize fixed effects to account for differing values across rounds. We estimated this regression separately for men and women. The coefficient on the DSG indicator, which describes the election effect, was 53.98 for men and 52.99 for women. Men had a slightly larger election effect, but this difference was small and insignificant. (Full table and results available; omitted for length).

F.3 Re-election Effects

We first assessed whether there is any re-election effect in the DSG. We regressed 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. We then added 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. (Results omitted because of journal length limits.)

F.4 Confidence

F.4.1 Confidence as measured by ICG payoffs

The main manuscript described how we replicated the main table using ICG payoffs as a measure of confidence. For space, we show that in Table F.4, which replicates Table 1. It shows average ICG payoffs for leaders versus non-leaders, and candidates versus non-candidates.

Here, we show statistical analysis of how confidence does not explain the selection patterns we found based on NMVW. We also show that our measure of NMVW still explains selection patterns, even when we control for the ICG payoffs as a measure of confidence.

Table F.5 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 insignificant.

In Column 2, we also include our NMVW measure based on average ICG ticket purchases, and in Column 3, we do the same for our measure based on estimates of Nash NMVW. 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 NMVW measures still have similar,

Table F.3: Differences Between Candidates and Not Candidates, by Gender

	Dependent variable:		
	Candidate		
	(1)	(2)	(3)
Female	0.817	0.849	0.892
	(0.139)	(0.137)	(0.138)
Risk Av. (LVR 1225)	0.002		
	(0.001)		
Risk Av. (LVR 1605)		0.001	
		(0.0005)	
Risk Av. (1225 1605)			0.116
			(0.157)
ICG NMV Ave.	0.007	0.008	0.007
	(0.008)	(0.008)	(0.008)
Female x RA (LVR 1225)	0.004		
	(0.003)		
Female X RA (LVR 1605)		0.002	
		(0.002)	
Female X RA (1225 1605)			0.444
			(0.270)
Female x ICG NMV Ave.	0.081	0.081	0.085
	(0.014)	(0.014)	(0.014)
Constant	0.381	0.366	0.309
	(0.079)	(0.079)	(0.086)
Observations	1,750	1,750	1,750
Log Likelihood	1,112.243	1,113.634	1,111.175
Akaike Inf. Crit.	2,236.485	2,239.269	2,234.350
Note:			p<0.1; p<0.05; p<0.01
			Averages given in 100s of tickets.

Table F.4: 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.

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 NMVW, persists.

We then replicated that analysis again focusing on the likelihood of winning an election, conditional on selecting into candidacy, in Table F.6. The dependent variable was a binary indicator for whether the candidate wins. As above, we could not reject the null that average ICG payoff has the same effect for men as for women. And as above, our NMVW measures still did have different effects on the probability of winning for men and women candidates.

Table F.5: 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)	
Nash NMVW			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 Nash NMVW			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
Note: Averages given in 100s of tickets.		p<0.1; p<0.05;	p<0.01

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 F.7 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 F.6: 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)	
Nash NMVW			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 Nash NMVW			0.061
			(0.027)
	(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: Averages given in 100s of tickets. $p<0.1$; $p<0.05$; $p<0.01$
Averages given in 100s of tickets.

Table F.7: 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
Note:	p<0.1; p<0.05; p<0.01	

F.4.2 Confidence as measured by electoral wins

A third aspect of confidence refers to one's beliefs about their ability to win elections. This facet of confidence 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 confidence with gender differences in election aversion. Though, Bernhard and de Benedictis-Kessner (2021) do not find that women candidates are especially discouraged by electoral losses.

Here, we return to the aspect of confidence that is specific to one's ability to win elections. Women and men definitely differ in these beliefs. However, this difference alone is insufficient to explain our results in the main manuscript. It would need to be that case that boosts to electoral confidence, presumably from winning elections, affected men and women differently. We do not find evidence of this.

Table F.8, 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.[44] We also interact these indicators with gender. Candidacy decisions tend to be persistent. Both men and women 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 hits) to their electoral confidence. Women candidates who lost were more persistent in their candidacy decisions. If anything, this suggests that the type of women who run are less influenced by in-game confidence effects. 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 find the confidence boosts or hits affect

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

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 F.8: Effect of Electoral Wins/Losses on Candidacy/Winning, by Gender

	Candidate	Electoral Success
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

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

Appendix Only 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. 2021. "Men and women candidates are similarly persistent after losing elections." *Proceedings of the National Academy of Sciences* 118(26).

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

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

Park, Yon Soo, Sarah Hummel and Stephen Chaudoin. 2022. "The Election Effect: Democracy in Inter-Group Contests.". Working paper, Harvard University.

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