

# Elections, War, and Gender: Self-Selection and the Pursuit of Victory

January 13, 2024

## **Abstract**

Why might female leaders of democratic countries commit more resources - money, equipment, soldiers, etc. - to interstate conflicts than male leaders? Existing literature suggests the desire to counter gender stereotypes helps explain this behavior. In this research note, we argue that the non-randomness of democratic election plays a complimentary role: since running for office is generally more costly for women than men, only women who value winning competitions more highly choose to become candidates for political office. These women also aggressively pursue victory in conflict situations that arise after their election. To provide microfoundational evidence for this claim, we used an online laboratory featuring real-time, group play in which participants chose to run for election, conducted a simple campaign, and represented their group in a contest game if elected. Women who placed a higher non-monetary value on winning were more likely to self-select into candidacy and, when victorious, spent more resources in intergroup contests than not only the average female participant spent in one-on-one contests, but also than male elected leaders.

*“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

There is a complex relationship between leader gender and behavior in interstate conflicts.<sup>1</sup> Gender bias and gender-based stereotypes create different strategic environments for male and female leaders. Female executives may feel pressured to combat negative stereotypes and prove their toughness, especially in the domains of interstate conflict and national security. In line with this, prominent studies find that female chief executives behave similarly to their male counterparts<sup>2</sup> or even pursue *more* hawkish foreign policies,<sup>3</sup> despite the fact that women in the general public espouse more pacifist opinions than men.<sup>4</sup>

We argue that gendered differences in the types of individuals that self-select into candidacy for political office also explain these behavioral patterns at the international level. Running for office is costlier for women than men, both materially and psychologically. Women are less likely to be recruited, may face perceived or real gender bias at the polls, and often place a lower relative value on holding office.<sup>5</sup> Existing research also documents greater competition aversion, on average, among women, leading them to avoid competitive situations like electoral contests, regardless of their actual quality or chance of success.<sup>6</sup> Faced with steeper costs to candidacy, only women with a strong intrinsic drive to win contests - a high “non-monetary value to winning” - choose to enter elections. This same trait subsequently affects choices in inter-group conflicts, leading them to dedicate more resources to the pursuit of victory than the average woman prefers and even - when the selection effect is strong enough - than does the average male elected leader.

---

<sup>1</sup>While we recognize the distinction between sex and gender, we follow the literature and use the term gender to describe leader attributes.

<sup>2</sup>E.g. Horowitz, Stam and Ellis (2015).

<sup>3</sup>E.g. Koch and Fulton (2011).

<sup>4</sup>E.g. Eichenberg (2019).

<sup>5</sup>Anzia and Berry (2011), Ashworth, Berry and Bueno de Mesquita (2023), Lawless and Fox (2010).

<sup>6</sup>E.g. Kanthak and Woon (2015).

To be clear, these selection dynamics do not preclude the possibility that gender stereotypes cause female leaders to act more hawkishly at the international level than they would otherwise. But our argument raises the possibility that at least some female leaders choose to dedicate more resources to defense policies and/or winning ongoing wars because they are predisposed to do so. Both self-selection and stereotype-based mechanisms may be at work, but to ignore the former is to over-inflate the estimated effect of the latter.

Our argument thus refocuses attention on the preferences and beliefs of female leaders themselves. Biographers of these individuals, including Margaret Thatcher, Indira Gandhi, and Golda Meir often point to their “unshakeable convictions.”<sup>7</sup> Such leaders are perhaps the least likely to be pushed into either conforming to stereotypes or engaging in costly counter-stereotype behavior, when doing so does not reflect their beliefs about appropriate conflict behavior. Even so, our argument does not negate the role of gender bias. The gap in costs to candidacy between men and women is greater in more sexist societies, strengthening the selection mechanism such that only women with more extreme preferences choose to run for office.

Identifying the effect of leader gender in observational studies is difficult because leader gender is not randomly assigned.<sup>8</sup> In a call for more micro-level evidence linking gender and conflict, Cohen and Karim (2021) note that endogeneity concerns make analyzing the theoretical mechanisms particularly difficult. The list of female leaders is also frustratingly short, which raises the possibility that specific leaders or situations have an out-sized influence on estimates.<sup>9</sup> Selection mechanisms are particularly difficult to isolate in observational studies because we rarely have comparable data about key features of people who do *and do not* eventually become leaders.

To address these concerns, we turn to a controlled laboratory setting. Participants played contest games in which they chose how much costly effort to expend in zero-sum competitions. Contest games have natural parallels to interstate conflicts, in which each side can spend more on defense and armaments to increase their chances of winning a war. Participants first played as

---

<sup>7</sup>Steinberg (2008) p 8.

<sup>8</sup>Judge (2021).

<sup>9</sup>Imamverdiyeva and Shea (2022).

individuals in one-on-one contests. They later played in groups that endogenously elected their leaders. Participants decided whether to run for election, wrote short campaign messages, and voted for their preferred candidates. The winning candidate played a contest game on behalf of their group against the leader of another group. Participants were anonymous throughout and groups were shuffled frequently, minimizing the role that gender stereotyping played in both the election and contest games.

Our results reproduce the observation motivating many studies of leader gender and conflict: female elected leaders chose to invest *more* resources in intergroup contests than male leaders, despite that fact that women, on average, chose to allocate fewer resources than men in individual contests. Consistent with our theoretical expectations, women who placed a higher non-monetary value on winning contests were significantly more likely to run for office and win than women who placed a lower non-monetary value on winning. This dynamic was much weaker for men. We also rule out selection based on alternative characteristics, notably risk preferences and confidence.

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. This is not possible in the real world since the gender of candidates and 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.<sup>10</sup> Following Renshon, Lee and Tingley (2017), we therefore view the evidence presented here as “part of larger bodies of evidence compiled from different data sources and different research designs” (S204).

We are not the first to speculate about the international implications of gendered selection dynamics. For instance, Enloe (1989) wondered whether female leaders are outliers with respect to conflict behavior who - in Margaret Thatcher’s words from above - “got through.” In a call for future research over 25 years later, Reiter (2015) likewise posits that the “conflict attitudes of

---

<sup>10</sup>Kertzer (Forthcoming).

women elected to office [may not] represent the conflict attitudes of all women” (1313). Our study answers these calls and provides direct evidence to support the importance of electoral selection for subsequent conflict behavior. By specifying the traits driving both selection and leader choices, we also develop a theoretical framework for better understanding the relationship between democratic electoral dynamics and patterns of international behavior.

## Gender, Self-Selection, and War

On average, women hold more pacifist preferences than men. A recent meta-analysis of 17 survey experiments finds that female respondents are less supportive of the use of force in every included study.<sup>11</sup> Analyzing over 900 American public opinion survey questions covering 24 use-of-force cases from 1982 to 2013, Eichenberg (2019) finds that women are almost always less supportive of using force, although there is important variation in effect size. This study also shows that the difference between men and women’s opinions about the use of force generally persists in a cross-national examination of 62 countries.<sup>12</sup> These findings underpin the claim that political leaders are more restrained after the expansion of women’s suffrage, since they have to account for the more pacifist preferences of female constituents.<sup>13</sup>

However, such findings do not necessarily imply that female *leaders* pursue more peaceful policies than their male counterparts. Indeed, some existing research finds that female leaders have higher levels of defense spending, are involved in more violent interstate disputes, and initiate more international conflicts.<sup>14</sup> While definitive evaluations of this claim are complicated by the paucity of female leaders in the data,<sup>15</sup> the evidence - at a minimum - seems to refute the claim that female leaders reflect the less hawkish preferences of the average woman.<sup>16</sup>

---

<sup>11</sup>Barnhart et al. (2020).

<sup>12</sup>For earlier work on US public opinion, see Shapiro and Mahajan (1986). For additional cross-country comparisons, see Jelen, Thomas and Wilcox (1994)

<sup>13</sup>Barnhart et al. (2020).

<sup>14</sup>Caprioli and Boyer (2001); Dube and Harish (2020); Koch and Fulton (2011); Powell and Mukazhanova-Powell (2019); Schramm and Stark (2020); Trager and Barnhart (2023).

<sup>15</sup>Horowitz, Stam and Ellis (2015).

<sup>16</sup>Burns and Bowling (2021); Imamverdiyeva and Shea (2022).

The most prominent explanations for the disconnect between conflict preferences of female citizens and the choices of female executives emphasize gender bias in how female leaders are treated. Female leaders must compensate for stereotypes of women as weak by choosing policies that counteract that stereotype. Pressure can come from voters, whose biases push women to “prove themselves.” For instance, Blair and Schwartz (2021) use survey experiments to show that respondents are less approving of conciliatory policies when taken by female, rather than male, leaders. Post and Sen (2020) argue that stereotypes held by opposing leaders also matter. 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, ultimately resulting in higher hostility levels. In short, female leaders behave more aggressively in conflict situations than they would choose to in the absence of a gendered strategic environment.

We propose an alternative mechanism that also explains the divergence between the preferences of female citizens and the actions of female leaders: the democratic leader selection process results in female leaders who place a higher value on winning than average and therefore invest more in winning any interstate conflicts that arise once elected. The two mechanisms are not mutually exclusive. Some female leaders may be acting on their true preferences, while others respond to stereotypes. Stereotypes might also push women who already value winning more highly than the general female population to behave *even more* aggressively in the pursuit of victory. In either case, ignoring the role of selection causes us to overestimate the effect of stereotypes on the behavior of female leaders.

Why might women who value winning highly be over-represented at the level of executive office? To answer this, we draw on the robust literature about gender and democratic elections. This literature points to a variety of institutional, cultural, and psychological reasons why running for office is less attractive, in general, for women than men.

Some studies emphasize psychological differences in how men and women perceive competitions. Niederle and Vesterlund (2007) and a large body of subsequent literature document a greater competition aversion among women, since “...the prospect of engaging in a future competition

may cause women to anticipate a psychic cost and deter them from tournaments (p. 1070).” At their core, elections are competitions: they are zero-sum contests that only one candidate can win. Laboratory research also finds that women are less likely to self-select into candidacy for election, even if they are equally or more qualified than men.<sup>17</sup> Likewise, 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. These psychological differences clearly impact how women and men decide whether or not to run for office, but they are far from the only influential factors. Women may also face higher material costs to candidacy. For example, women are less likely to be recruited, placing a greater onus on them to proactively seek out electoral opportunities.<sup>18</sup> Voters may be less likely to vote for female candidates than male candidates of the same quality.<sup>19</sup> The net value of office may also be lower for women than men. Women continue to face greater competing demands - notably, familial and professional - on their time,<sup>20</sup> meaning that they must give up more to hold office. Concerns about gender bias in politics may also make holding office less attractive to women.

These factors combine to make running for office systematically more costly for women. Lower rates of candidacy among women in the real world reflect this fact. It also implies that women who overcome these higher candidacy costs may systematically differ from those who do not.

A second theoretical component, also from experimental literature, helps us understand this difference. 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.<sup>21</sup> We call this one’s non-monetary value to winning (henceforth, “NMVW”). Although our terminology describes this as a value to winning, this concept can also capture any additional sting from losing. Unlike many factors that affect the desirability of becoming an electoral candidate, the NMVW influences both the decision to enter an electoral contest and how individuals behave after winning. A higher

---

<sup>17</sup>Kanthak and Woon (2015).

<sup>18</sup>Lawless and Fox (2010).

<sup>19</sup>Anzia and Berry (2011), Ashworth, Berry and Bueno de Mesquita (2023).

<sup>20</sup>Lawless and Fox (2010).

<sup>21</sup>Sheremeta (2010).

NMVW both makes entering an electoral contest more attractive *and* increases the marginal benefit of spending more or exerting greater effort to ensure victory in subsequent contests.

Together, the gender gap in candidacy costs and heterogeneity in NMVWs explain why female leaders may invest more resources in pursuing victory than their male counterparts, despite women being more pacifist in general. Women require a higher NMVW than men to enter a competition, especially when the gap in candidacy costs is large. This screens out women with low NMVWs from leadership. Women with high NMVWs also devote more resources to win competitions and conflicts after their election.

The strength of the selection mechanism depends on how much more costly entering an election is for women than men. On average, men face lower barriers to entry, resulting in a pool of males running for office with a wider range of NMVWs than the pool of potential female leaders. However, the gap between these pools depends on societal context.

At one extreme are societies where politics is so male-dominated that even women with exceptionally high NMVWs rarely choose to run and, if they do, are not elected. In such societies, potential female leaders require more than just a high NMVW to obtain office. For instance, Baturo and Gray (2018) demonstrate that female leaders are particularly likely to be part of a political family in contexts where women's suffrage is relatively new; that is, before female participation in politics becomes normalized. Even so, many wives and daughters of male political leaders do not pursue office. Political connections make running more attractive - or, perhaps, possible - but it still takes a high NMVW to overcome the barriers to entry in male-dominated societies.

On the other hand, existing work argues that societal equality and female empowerment has a direct, pacifying, effect on interstate conflict behavior.<sup>22</sup> We argue that societal equality also moderates the relationship between leader gender and conflict by weakening the degree to which leader selection mechanisms are gendered. In more equal societies, the candidacy considerations faced by women more closely resemble those faced by men, causing more women to run for political office, and shrinking the observed difference in leader behavior across genders, at the

---

<sup>22</sup>Cohen and Karim (2021).

international level. Consistent with this, Koch and Fulton (2011) find that the percent of women in legislatures moderates the (positive) effect of female chief executives on defense spending. Where a gender gap in candidacy costs exists, female leaders invest more in winning conflicts than would the average female citizen. Where the gender gap in candidacy cost is large enough, they will be even more hawkish than male leaders.

Notably, this selection mechanism helps explain the choices female leaders make when preparing for future conflicts and their investment in ongoing ones. Since the likelihood of winning a conflict depends, at least in part, on the level of preparation, we expect female leaders to engage in greater defense spending in times of both war and peace than the average female citizen would prefer. In line with this, Imamverdiyeva and Shea (2022) finds that female executives spent roughly the same on defense as carefully constructed synthetic (male) control observations, while Koch and Fulton (2011) find that countries with female executives actually have greater defense spending than those with male executives. There is also evidence that female leaders fight harder than male leaders once a conflict begins. Post and Sen (2020) show that, conditional on initiating a crisis, hostility levels are higher for women than men, and Caprioli and Boyer (2001) find the severity of violence within a crisis is greater for states with female chief executives.

A corollary to this is that women with higher NMVWs may also be more likely to win elections. Just as they invest more in winning inter-group conflicts once elected, a higher NMVW results in greater effort spent in pursuit of electoral victory. However, we do not want to overstate this. An individual's NMVW does not, in itself, affect their attractiveness to voters in a predictable way. Furthermore, while some elections may be won or lost based on candidates' efforts, this may be marginal in higher stakes races.

We combine these observations to formulate a general hypothesis:

**Hypothesis.** *Individuals with higher non-monetary values to winning contests are more likely to run for office, win elections, and exert more costly effort to win conflicts while in office. This relationship is stronger for women than men.*

While the selection mechanism has clear implications for how leaders prepare for and fight

ongoing conflicts, expectations about conflict initiation are less straightforward. The same is true for stereotype-based mechanisms. For example, Reiter and Wolford (2022) show that incorporating a private benefit from defeating a female-led opponent, or a private cost to losing to one, can either increase or decrease the probability of starting a war. Schwartz and Blair (2020) show that survey respondents more harshly punish female leaders for backing down from threats. This could reduce conflict by facilitating stronger signals of resolve from female leaders; it could also encourage adventurism and make it harder for a female leader to de-escalate (p 890).

Our selection mechanism has similarly complex implications for conflict initiation. Some of the factors that generate the difference in candidacy costs between men and women - e.g. competition aversion - also make conflict initiation *less* appealing to female leaders than men. However, selection at the electoral stage implies that female leaders have, on average, higher NMVWs than male leaders. This makes initiation *more* attractive to female leaders by increasing both the value of victory and the likelihood they ultimately win. It is beyond the scope of this note to analyze crisis initiation, but we formalize the intuition in Appendix A. Roughly, whether the gender gap at the electoral level - which alters the distribution of NMVWs among female leaders - or the gender gap at the international level is larger will determine whether men or women initiate conflict at a higher rate. Importantly, this analysis also demonstrates that the predictions about selection and effort in an ongoing conflicts do not depend on whether crisis initiation is exogenous or endogenous.

To the best of our knowledge, the closest related observational study to ours is Imamverdiyeva and Shea (2022). They use synthetic controls to demonstrate that male and female leaders have similar levels of military spending, which they argue is evidence of non-random selection and/or stereotype-driven decisions by women. They then examine lame duck female leaders who are arguably less constrained by voter stereotypes. While the number of cases is small, they conclude that they cannot rule out the possibility that selection contributes to the similarity between male and female leaders. Our research leverages the laboratory setting to provide more definitive evidence of selection's role.

## Empirical Design

We recruited 162 participants for 10 sessions of our game from Amazon’s Mechanical Turk (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.<sup>23</sup> As in war, players decide how much to spend in pursuit of a zero-sum prize. In each round, players started with an endowment of 1000 points that they could spend to buy up to 1000 lottery tickets. Players kept any points they did not spend on tickets. Each player’s likelihood of winning the contest equaled the number of tickets they bought divided by the total number of tickets bought. The value of the prize varied each round, generally from 1225 to 2715 points. The winner of the contest received the prize.<sup>24</sup>

At the start of each session, we randomly paired respondents in each of 12 rounds to play one-on-one contests. Participants were not identified and pairs were re-shuffled every round. 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.”<sup>25</sup> Group members remained anonymous, without any gender identifiers. We re-shuffled 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 wrote short campaign messages that were shown to their group members. Participants then voted for their group leader.<sup>26</sup> The winning candidate received a 245 point bonus, which could not be used in the contest. This protocol reflects two key features of real-world elections: (1) participants must *self-*

---

<sup>23</sup>Chaudoin and Woon (2018).

<sup>24</sup>Appendix B provides more detail about the protocol, sample, and compensation. Participants watched a video explaining the rules before playing, produced by a graphics designer to make it clear and engaging. We paid all participants \$5 plus \$1 for every 210 points they won in the average of 5 randomly selected rounds, calculated at the end of their session. Players did not accrue points over rounds. Participants knew this and were encouraged to think of every round as a separate decision task.

<sup>25</sup>We told participants each group had 7 members. However, groups started with 8 members, as a safeguard against dropouts. In practice, dropouts were very rare. Appendix C provides a detailed justification for this deception.

<sup>26</sup>Candidates could vote for themselves. Ties were broken randomly.

*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 group won the contest, each group member, including the leader, received the prize amount. In other words, the group won or lost together and leaders could not discriminate between group members for either ticket purchases or the distribution of winnings. In terms of expected utility, the leaders' decisions in the DSG were strategically identical to those made by individuals in the ICG.

Comparing the ICG and DSG helps us overcome an important obstacle in studying selection: the need to observe the decisions of participants that do *and do not* eventually become leaders. While some innovative research designs address this challenge<sup>27</sup>, it is not practical in most observational studies. Participants' ICG behavior provides data about how each they 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.

We purposefully ensured participant gender was unknown throughout. While this is an obvious abstraction from reality, it helps us isolate the effect of selection as much as possible. Of course, even understanding that one's gender will not be explicitly revealed does not entirely remove all effects of gender bias. Internalized stereotypes might lead female leaders to act as though they have something to prove, even if they are aware their gender is unknown to others. Given the deep socialization of gender norms, there is no way to remove this entirely. However, the anonymous laboratory setting allows us to get as close as possible to this ideal, rendering female participants "freer" to make choices that reflect their actual preferences.

---

<sup>27</sup>E.g. Bernhard, Shames and Teele (2020).

## Measuring NMVW

We use two measures of 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:  $T = \frac{p+v}{4}$ . An estimate of the NMVW is therefore given by  $v = 4T - p$ .<sup>28</sup> By participant, we calculate the average value of  $v$  from the ICG rounds and call this their “Nash NMVW.”

Second, we use the average number of tickets purchased in the ICG rounds. An individual with a higher NMVW purchases more tickets than an individual with a lower NMVW. This measure thus reflects the quantity of interest, even it does not directly measure it, and has the advantage of not using Nash Equilibrium strategies as a benchmark. In practice, all the results are very similar for both measures.

## Empirical Results

### Gender Differences in Leader Behavior

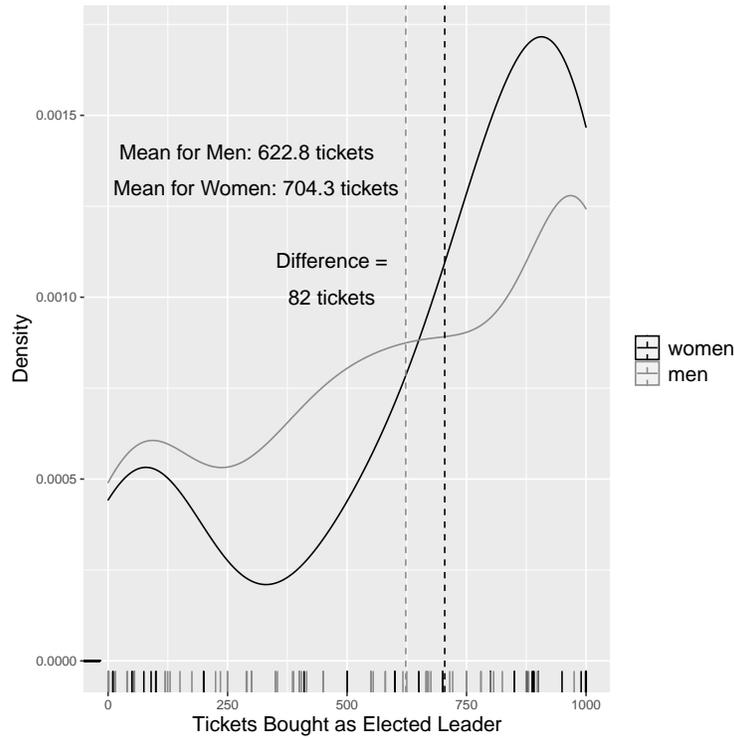
In our data, elected female leaders spent substantially more on contests than male leaders. Figure 1 shows the number of tickets elected leaders bought in intergroup contests, split by gender. Female elected leaders purchased, on average, 82 more tickets (from a maximum possible 1000 tickets) per round than male leaders. This is a large spending increase of over 8% of the total budget available. This difference 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.<sup>29</sup>

---

<sup>28</sup>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 yield the Nash prediction.

<sup>29</sup>See Appendix D.

Figure 1: Distribution of Tickets Bought in the DSG, by Gender



Note: The lines show the smoothed density of the distribution of tickets bought by elected women (black) and men (gray). Vertical lines show the mean of tickets bought by gender

### Evidence of the Selection Effect

The difference in ICG behavior for leaders and non-leaders is much bigger for women than men. This aligns with our expectation that only women who value winning highly choose to run in elections. 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.<sup>30</sup> Among men, there are positive, but small, differences between leaders and non-leaders. Male leaders had a Nash NMVW that was 32 points higher and they purchased 8 more tickets per round, on average, in the ICG, compared to male non-leaders. The differences are much larger for women. Female leaders had a Nash NMVW that was 598 points higher and purchased 150 more tickets, on average, in the

<sup>30</sup>A participant's ICG behavior "counts" toward the leader mean in rounds that s/he is a leader. This accounts for how some participants became leaders more often by weighting the overall means accordingly.

ICG compared to female non-leaders. In short the degree of sorting is starkly different for men and women.

Table 1: Differences in NMVWs, by Gender

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Ave. Nash NMVW	539	507	958	360
<i>difference</i>		+32		+598
Ave. ICG Tickets	565	557	670	520
<i>difference</i>		+8		+150

<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Ave. Nash NMVW	520	506	741	259
<i>difference</i>		+14		+482
Ave. ICG Tickets	560	557	616	495
<i>difference</i>		+3		+121

<i>(III) Winning vs Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
Ave. Nash NMVW	522	508	979	662
<i>difference</i>		+14		+317
Ave. ICG Tickets	561	557	675	596
<i>difference</i>		+4		+79

Note: Part III excludes candidates who ran unopposed.

Self-selection into candidacy explains the largest part of the gap between the NMVWs of female leaders and non-leaders, as shown by comparing the behavior of candidates and non-

candidates in the middle portion of Table 1. Among men, there are only 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 (560 vs. 557), on average, in the ICG, compared to male non-candidates. Among women, however, the differences are again striking. Women who ran for elections had an average Nash NMVW that was 482 points higher and purchased an average of 121 more tickets in the ICG (616 vs. 495), on average, compared to women who did not choose to run. 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). While election-winning men were very similar to losing male 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 during elections, this difference in success is attributable to differences in campaigns, analyzed more below.

These differences are statistically significant. 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. In columns 3-4, we do the same using an indicator for whether a participant self-selected into candidacy. In columns 5-6, we do the same using an indicator for whether a candidate won election. In each model, the interaction term is substantively and statistically significant: NMVW has a stronger positive effect on the probability of being a leader, candidate, and winner for women than men.<sup>31</sup> Figure 2 shows the estimated effects visually.

---

<sup>31</sup>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 E.

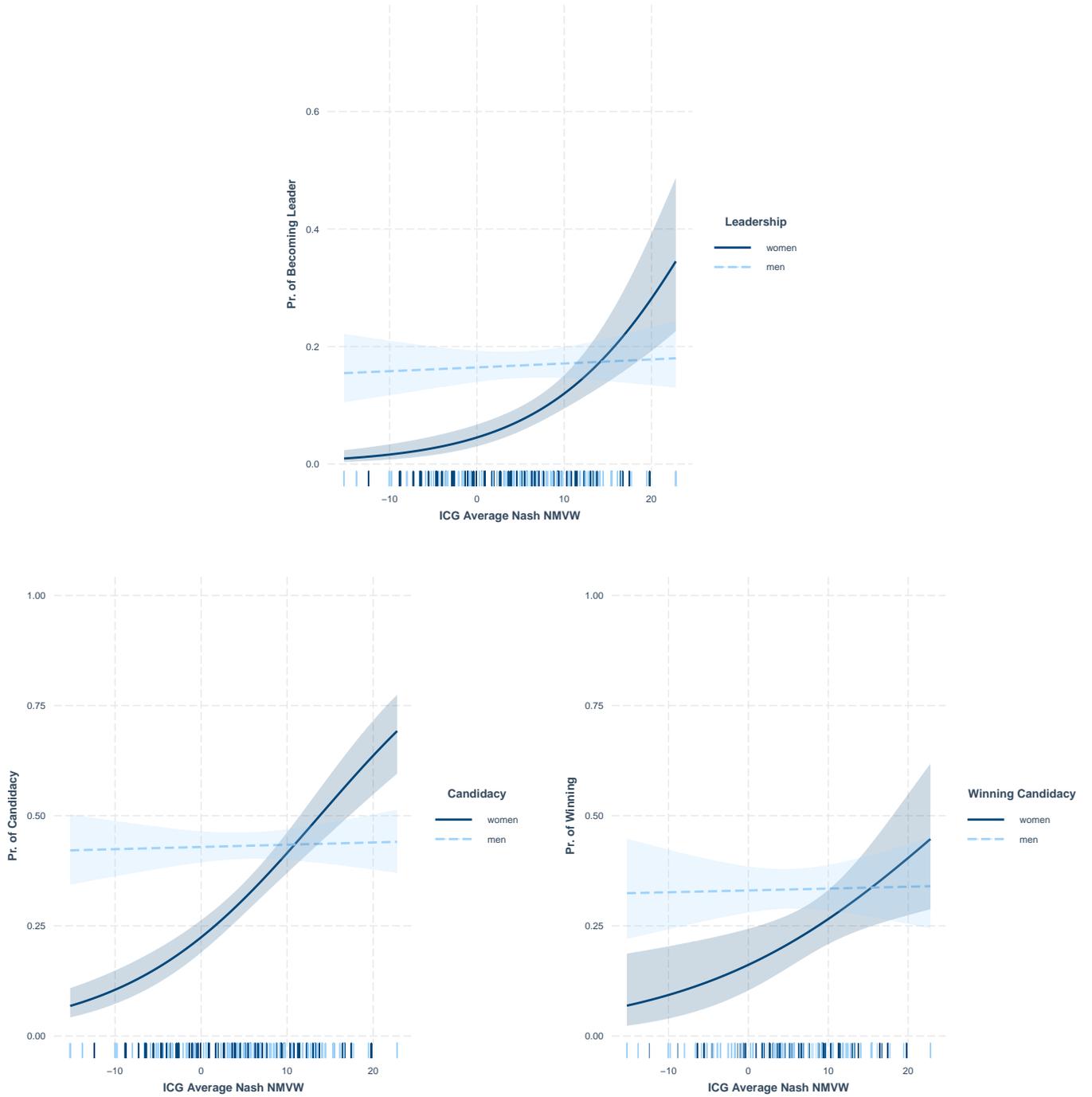
Table 2: Effect of NMVW on Leadership and Candidacy, by Gender

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

Figure 2: Predicted Effects of Nash NMVW, by Gender



Note: Plots show predicted probability of being a leader/candidate in a DSG round, using estimates from Table 2, columns 1, 3 & 5.

The results are consistent and striking. Only women with the the highest NMVWs run for and win office at rates comparable to men. They then proceed to invest more resources in inter-group competitions.

### **Gender and Campaign Strategies**

Selection into candidacy by high-NMVW women explains a large portion of the differences in leader behavior, but higher-NMVW women were also more likely to win elections. We posit that women with higher NMVWs invest more resources - here, time and effort - at the campaign stage and subsequently are more likely to be elected. In the real world, effort may only have a marginal effect on election outcomes. However, this finding provides further evidence that our NMVW measures capture meaningful differences across individuals, with multiple implications for observable behavior.

Campaign message length is a reasonable proxy for effort.<sup>32</sup> Writing longer messages takes additional time and cognitive energy. Mturk participants prioritize task speed in order to maximize their wages per hour. Browser extensions help Turkers scrape for higher-paying, shorter tasks, with links to reviews showing the task's pay per hour. Unfortunately, we do not have a measure of the time spent entering their message.

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. Importantly, longer messages were more successful. An additional 50 characters raised a candidate's probability of winning by approximately 13%.<sup>33</sup>

While length is a reasonable proxy for effort, we would ideally measure message *quality* or *persuasiveness*, which likely increase with the effort put into crafting a message. This is subjective and difficult. It is further complicated since men and women, and participants with different NMVWs, chose different types of electoral appeals. Furthermore, messages with similar content

---

<sup>32</sup>Giovannoni and Feltovich (2022).

<sup>33</sup>See Appendix F. Mediation analysis also demonstrates that higher NMVW women wrote longer messages, which helped them win.

Table 3: Length of Campaign Messages, by Effort and Gender

Female	-38.598*** (10.531)	-16.420*** (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*** (5.248)	69.352*** (2.159)
Observations	676	676

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

can vary in their persuasiveness.

In Appendix F, we demonstrate that higher-NMVW women likely crafted higher quality messages – even when accounting for message length and content. This pattern does not appear for men. We coded whether each message fit into ten qualitative categories. We then regressed whether a person won election on their gender and NMVW, controlling for message length and content category. Among women, a higher NMVW was still associated with a greater probability of winning. This suggests that higher NMVW women did not simply write certain message types; they wrote *better* messages within these categories. Likewise, they not only wrote longer messages than low NMVW women; they also wrote better ones. These findings are much weaker for men.

## Alternative Explanations

### Risk aversion

An alternative explanation is that selection on risk preferences, not NMVW, influences behavior in post-election contests. It could be the case that only the most risk tolerant women select into candidacy and subsequently purchase additional tickets. To explore this, we leverage the fact that risk aversion affects ticket purchases differently as the prize value changes. As the prize value increases, the *marginal* increase in ticket purchases should be higher among more risk tolerant participants.<sup>34</sup> The intuition is straightforward: risk acceptant participants are more enticed to gamble as the prize increases than risk averse participants.

In the ICG, we included two rounds in which 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 275/280 points and their average ticket purchase when the prize was 1225/1235. We then calculated their average percent increase in ticket purchases as they moved from the low to high value. We use this as a measure of risk preferences across participants.

---

<sup>34</sup>Note that the relationship between tickets purchased and risk preferences is complex. Empirically, more risk averse participants purchase fewer tickets.

Higher values indicating greater risk acceptance.

Table 4 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 7 times larger than their purchase in the low value rounds. 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 and winning versus losing candidates: Women candidates (winning candidates) increased their ticket purchases to a smaller degree than those who were not candidates (losing candidates).

Table 4: Differences in ICG behavior, Risk Aversion (LVR and 1225 rounds), by Gender

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Av. Perc change in tickets	35.75	20.49	7.05	15.03
<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Av. Perc change in tickets	28.91	19.51	10.43	16.96
<i>(III) Winning vs. Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
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 compare the 275/280 rounds with the 1225/1235 values because these are the next highest values. Also, those prize values are only marginally above 1000-point ticket purchase cap. At higher prize values, the cap is more likely to bind, making it harder to assess participants' risk preferences. However, our results - and their inconsistency with explanations based on risk aver-

sion - obtain when we compare other pairs of prize values.<sup>35</sup> The Appendix also shows that our main results - about the differential effect of NMVW for women - obtain, even when controlling for these risk aversion measures.

### **Confidence and Learning**

It is also possible that participants self-select based on confidence, which could differ by gender. The underlying task in our game is strategic - payoffs depends both players' choices – and payoffs have a stochastic element. Players might gain confidence as they play - by observing their own payoffs and/or by learning about the play of others - in ways that could differ across genders.

In the first instance, receiving more points gives participants positive information about their performance relative to others. Winning more points in the ICG could increase a participant's confidence in their decision-making. We calculated the average ICG payoff for each player. Leaders in the DSG tended to have higher ICG payoffs than non-leaders. However, these differences are slightly larger for men (1401 vs. 1308) than women (1377 vs. 1310). The pattern is similar for candidates versus non-candidates and winning versus losing candidates. Any selection effects based on ICG payoffs were thus weaker for women than men.

In the second instance, participants could gain confidence as they learn about the behavior of others. Generally, players in contest games decrease ticket purchases over time,<sup>36</sup> which is usually attributed to learning. It could be that eventual female leaders did not adjust their purchases downwards over time to the same degree as men and women who were never leaders. This was not the case. Trends in ticket purchases as the ICG progressed were generally the same across men and women, and did not vary systematically by candidate or leader status. Taken together, this is evidence that selection was not driven by confidence or learning.<sup>37</sup>

---

<sup>35</sup>See appendix.

<sup>36</sup>Sheremeta (2010).

<sup>37</sup>Appendix G.

## Conclusion

To lead a democratic country, one must first win office. The electoral gauntlet is not for the faint of heart. Women face particularly high costs to candidacy. As a result, the very attributes that make women more likely to run for office also make female leaders more willing to spend resources on subsequent contests and conflicts. We demonstrated that female elected leaders invested more in costly contests than their male counterparts in a controlled laboratory setting. Women who chose to run and then eventually won elections most highly valued winning for winning's sake. The selection effect was not nearly as stark for men. These gendered patterns despite our minimizing the effects of explicit gender-bias and stereotyping by anonymizing participants.

The implication is that we should not attribute the decisions of female leaders solely to the constraints placed on them by their environment. While sexism, stereotypes – as well as gender-neutral strategic considerations – undoubtedly shape the choices of female leaders, their decisions also reflect held and/or learned preferences and beliefs. Writing about Meir, Gandhi, and Thatcher, 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.

While scholars have speculated that selection mechanisms play a role in the choices of female leaders, our research establishes micro-foundations for these arguments. We go beyond saying that female leaders are not representative of the general population, specifying the dimensions that are driving both non-random selection and leader decisions. We identify differences in psychological and material candidacy costs across genders as the driving force behind non-random selection. If this difference declines with time, then we may see female leaders whose actions more closely align with the preferences of their (female) constituents. Until then, female leaders will continue to be outliers with respect to conflict preferences and behavior.

We see at least two avenues for future research. First, we would like to examine the sensitivity

of leader behavior to changes in the gender gap in candidacy costs at the domestic level. As discussed above, there is some observational data that female representation moderates the behavior of female executives. A laboratory study would allow us to identify the mechanism behind this observation. For instance, following Preece and Stoddard (2015), we could prime candidates about the competitive nature of elections. If this increases the cost of candidacy more among women than men, it could exacerbate the gender gap at the inter-group contest stage.

Second, we focused on electoral selection. However, elections are not the only channels through which individuals obtain positions of influence. The most obvious comparison is with authoritarian leaders, where we anticipate the combination of (often) patriarchal norms and particularly fierce competition for leadership positions to further strengthen selection effects. Even within democracies, however, there is interesting variation in selection procedures. For example, Bashevkin (2018) examines foreign policy leaders, including US Secretaries of State Hillary Clinton and Condoleezza Rice, noting that 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” (2).<sup>38</sup> Appointment to such positions may be partially shielded from public scrutiny and electoral biases. But there is also a competitive gauntlet, as candidates jockey for appointments. As with elections, there may be a gender gap in how costly it is to put oneself forward for prominent positions. Barnes and O’Brien (2018) demonstrate that internal politics and the external security situation affect the likelihood of women being appointed to defense ministries. Future work could help uncover whether non-random selection also affects the behavior of those holding key non-elected positions.

---

<sup>38</sup>See also Burns and Bowling (2021).

## 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.
- Ashworth, Scott, Christopher R Berry and Ethan Bueno de Mesquita. 2023. "Modeling Theories of Women's Underrepresentation in Elections." *American Journal of Political Science* .
- 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.
- 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. 2019. *Gender, War, and World Order: A Study of Public Opinion*. Cornell Studies in Security Affairs Cornell University Press.
- Enloe, Cynthia. 1989. *Bananas, beaches and bases: Making feminist sense of international politics*. Univ of California Press.
- 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.
- Jelen, Ted, Sue Thomas and Clyde Wilcox. 1994. "The Gender Gap in Comparative Perspective." *European Journal of Political Research* 25:171–186.
- 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 and Richard Fox. 2010. *It Still Takes a Candidate*. Cambridge University Press.
- 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.
- 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. "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.
- Shapiro, Robert and Harpreet Mahajan. 1986. "Gender Differences in Policy Preferences: A Summary of Trends from the 1960s to the 1980s." *The Public Opinion Quarterly* 50(1):42–61.
- 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.
- 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”**

**Appendix A** Formalization and Simulations (pp 2-15)

**Appendix B:** Protocol, Recruitment, Demographics (pp 16 - 20)

**Appendix C:** Deception Justification (pp 20-22)

**Appendix D:** Differences in Leader Behavior by Gender (pp 22-24)

**Appendix E:** Robustness of Effect of NMVW (pp 24-26)

**Appendix F:** Campaign Messages Analysis (pp 26-29)

**Appendix G:** Alternative Explanations (pp 30-47)

**Appendix References** (pp 48)

## A Appendix: Formalization and Simulations

This appendix formalizes how differential costs to entering contests and the non-monetary value to winning (NMVW) influence differences between men and women in:

1. Whether to run for election.
2. Whether to initiate an inter-group conflict as an elected leader.
3. How hard to fight in the conflict (i.e. resources spend to increase the chances of victory).

We emphasize how two things in the model affect each decision. First, we focus on the key argument from the main paper: the costs for entering an election tend to be higher among women, which - through non-random selection - leads to differences in how hard leaders fight. For simplicity, we focus on individual-specific fixed costs that must be paid to enter an election or inter-group conflict. The distribution of these costs differs by gender, with women drawing higher costs - on average - than men.

As discussed in the paper, this way of modelling of costs is broad and general. These costs could be psychological, as is the case if women are more competition-averse than men. The costs could also be based on party or electoral politics, as is the case if women have to work harder to convince their party to back them. In addition, other gendered aspects of electoral selection are isomorphic to gender differences in the costs of entering a contest. For example, if perceptions of the probability of winning an election vary by gender and/or the material value to winning office varies by gender, this will have the same effect as differential costs. These potential differences also mean that only higher-NMVW women will choose to run in an election, which in turn, means that elected women will fight harder than elected males, on average.

In line with the theoretical argument in the main manuscript, as the gap in the distribution of these costs for women versus men widens, this decreases the likelihood that women run for election, relative to men. Only women with higher NMVWs will choose to run for election. This, in turn, increases how hard elected women fight in inter-group conflicts, relative to elected males.

Second, we incorporate endogenous interstate conflict initiation into the model and we analyze how differential costs affect the likelihood of conflict initiation for men and women. Incorporating conflict initiation lets us show how the main arguments from the manuscript - about elections and effort in a contest - are the same, regardless of whether interstate conflict initiation is endogenous or exogenous (as it was in the laboratory protocol).

We can also show how the effect of gender on conflict initiation is complicated. Whether men or women are more likely to initiate conflict depends on the assumptions we make about gendered cost differentials at the domestic versus international levels. If women have higher costs to entering contests than men, this has two competing effects on whether men or women will be more likely to initiate interstate conflicts. First, there is a direct effect, whereby the same costs that discourage a woman from entering an electoral contest also make an interstate conflict less attractive. This pushes interstate conflict initiation rates among women downwards, relative to men. However, the selection mechanism that we have highlighted works in the opposite direction. If only the highest NMVW women enter electoral contests, then initiating an interstate conflict is relatively more attractive for women. They know that they will fight harder in these conflicts and be more likely to win.

Which effect dominates depends on whether the gendered cost differential is greater at the domestic level (for elections) or at the international level (for interstate conflicts). When we let gender dynamics vary in different ways at the domestic versus international level, we can show that the presence of differential costs make elected females more likely to initiate an inter-group conflict for some values of the parameter space and less likely for other values. As the size of the domestic difference in candidacy costs decreases relative to the difference at the international level, then the difference between male and female leaders declines and female leaders become less likely to initiate conflict relative to male leaders.

This is notable because this finding fits with existing work showing how societal equality attenuates the relationship between gender and conflict. For example, Koch and Fulton (2011) find that “As women gain greater access to politics, the need for female chief executives to prove themselves or overcome stereotypes against them [diminishes]” (5). Our model shows that this moderation result can also be driven by a blunting of selection effects. As domestic politics becomes more equal, there is less stark sorting among leaders, which leads to a convergence in the observed choices of male and female leaders.

## A.1 Model

In our laboratory protocol, individuals choose whether to run for an election. Elected leaders then choose how to behave in a subsequent inter-group contest. Our model follows a similar structure, but we also allow elected leaders to endogenously decide whether to initiate conflict with the opposing elected leader. After election, two leaders are randomly paired. If either chooses to initiate conflict, then both choose how hard to fight in that conflict. We treat the decision over whether to run for election and whether to initiate a conflict as separate decisions. In other words, potential candidates decide whether to run for election without considering their potential future payoffs from an inter-group conflict.

### A.1.1 Common Costs and Benefits

The following describes the parameters that are common to all players, regardless of gender. Without loss of generality, assume that the winner of the election gets a private benefit for winning,  $w > 0$ , and that the winner of a conflict also gets a benefit, that same  $w$ .<sup>39</sup> These benefits to winning elections/conflicts are common across everyone.

Choosing to run in the election and choosing to initiate a conflict both entail paying a fixed cost,  $k > 0$ , which is subtracted from the person’s payoffs regardless of whether they win the election or the conflict.<sup>40</sup> This is a fixed cost component that is also common across everyone. We normalize the payoffs to not entering an election to zero.

Each person draws a NMVW value,  $v \sim U[0, \bar{v}]$ .  $v$  is strictly positive, since this represents any additional benefit someone enjoys when they win a contest. The assumption of the functional form of the distribution doesn’t affect our arguments.

---

<sup>39</sup>Later, we return to the possibility that the stakes of an interstate conflict can be greater than those of an election. Relaxing this assumption does not change our results.

<sup>40</sup>Again, it does not change results to let these costs vary.

### A.1.2 Costs of Entering Election/Conflict and Gender

We assume there is a population of size  $N$ , with half women and half men. Each individual also draws a  $c$  value, which represents their *additional* personal cost to participating in an election or in an inter-group conflict. These costs are also strictly positive. The distribution of costs can differ for men and women. For men,  $c \sim U[\underline{c}_M, \bar{c}_M]$ , and  $\underline{c}_M > 0$ .

For women,  $c \sim U[\underline{c}_M + z, \bar{c}_M + z]$ , where  $z > 0$ . This is a simple way of specifying distributions so that the cost distribution for women first order stochastically dominates the cost distribution for men.<sup>41</sup>

### A.1.3 Election Choice

After each participant draws their personal  $c$  and  $v$  values, they decide whether to run for election. We assume that the probability that any individual wins election is exogenously given,  $p \in (0, 1)$ .<sup>42</sup> For simplicity, we assume that the probability that a participant wins if they choose to run is 0.5. This is equivalent to assuming that a participant's skill as a candidate is not affected by their NMVW or costs to running. Note, none of our arguments would change if we relaxed this assumption. An individual enters the election if her expected utility from doing so is greater than or equal to zero:

$$p(w + v_i) - k - c_i \geq 0 \quad (1)$$

The potential candidate therefore must draw a  $v$  value that is high enough, or draw a  $c$  value that is low enough, to to make running worthwhile.

### A.1.4 Effort in Inter-group Conflict

As in the main paper, we think of inter-group conflicts as contests, where the elected leader from both sides choose how much costly effort to exert to try and win the prize. We denote effort levels with  $e$ , which must be greater than or equal to zero. The probability that side  $i$  wins is a function of their effort level and that of their opponent  $j$ . We use the Tullock contest success function:  $p_i(e_i, e_j) = \frac{e_i}{e_i + e_j}$ . We set the marginal cost to effort at 1. This results in the following utility function:

$$EU_i(e_i | v_i, v_j) = \frac{e_i}{e_i + e_j(v_i, v_j)}(w + v_i) - e_i - c_i \quad (2)$$

---

<sup>41</sup>As above, the functional form of the distribution does not affect results.

<sup>42</sup>The assumption that the probability of winning is exogenous is identical to assuming that neither  $c$  nor  $v$  make a candidate more or less attractive in the election. This is different from citizen candidate models where citizens differ in their "quality" as a candidate, as in Ashworth, Berry and Bueno de Mesquita (2023). Note, that relaxing this assumption of an exogenous probability of winning would *not* change results. The natural way to endogenize the probability to winning would be to let the probability of winning election be increasing in  $v$ . As Ashworth, Berry and Bueno de Mesquita (2023) and others show, this results in cut-point based strategies that retain the same type of self-selection that we describe here.

For a player  $i$  who drew a value to winning  $v_i$ , matched against an opponent who drew a value of  $v_j$ ,  $i$ 's optimal effort level in an inter-group conflict, in equilibrium is:  $e_i^* = \frac{(w+v_i)^2(w+v_j)}{(v_i+v_j+2w)^2}$ .  $i$ 's optimal effort level is increasing in their own value to winning the prize ( $w + v_i$ ), and it also takes into account their opponent's values, which influence their opponent's effort choices.<sup>43</sup>

We simplify our analysis by assuming that an individual chooses their optimal effort level by best responding to an individual with a gender-specific average valuation. By the time individuals decide their effort levels in conflict, they know both the gender of their opponent and that their opponent chose to run for election.<sup>44</sup> The leader therefore knows that their opponent's draw of  $v$  was "high enough," relative to their draw of  $c$ , to justify running in the election. As mentioned above, Unfortunately, incomplete information contest games generally do *not* have closed form solutions for effort levels in equilibrium. Fey (2008), however, shows that the optimal effort levels (solved for numerically) in a game with continuous, uniformly distributed costs are "qualitatively similar" to the optimal effort levels in the full-information game. Optimal effort levels in the incomplete information game are essentially equal to optimal effort levels in the complete information game, just shifted downwards.<sup>45</sup> Therefore, we use the solution to the complete information game to approximate the solution to the incomplete information game, focusing on the equilibrium prediction for individual with average  $v_j$ , conditional on gender. We denote the mean valuation ( $v_i$ ) for men and women who chose to run with  $\bar{v}_M$  and  $\bar{v}_F$ , respectively.<sup>46</sup>

The best response function for player  $i$  when facing an opponent  $j$  with average valuation by gender:

$$e_i^* = \frac{(w + v_i)^2(w + \bar{v}_M)}{(v_i + \bar{v}_M + 2w)^2} \quad \text{Facing a male opponent}$$

$$e_i^* = \frac{(w + v_i)^2(w + \bar{v}_F)}{(v_i + \bar{v}_F + 2w)^2} \quad \text{Facing a female opponent}$$

Using these equations, we can write the expected effort level of a randomly drawn player, taking into account their own gender and the gender of their opponent. We denote  $i$ 's gender with the first letter of the subscript and  $j$ 's gender with the second letter. For example,  $\bar{e}_{MF}^*$  denotes "the expected equilibrium effort level of a male leader if he was facing a randomly chosen female opponent." In other words, these quantities denote how much effort the "average" male/female leader would exert in a conflict, if they were facing the "average" male/female opponent. From  $i$ 's perspective, these quantities matter because  $i$  needs to think about how hard  $j$  will fight. These quantities let  $i$  answer the question: "If I was paired with a random opponent, what's my best guess at how hard they will fight, taking into account their gender and also taking into account that they know my gender (but not my specific draws)?"

<sup>43</sup>See Chaudoin and Woon (2018).

<sup>44</sup>In the laboratory protocol, a leader does not know their opponent's gender. In this model, we assume that leaders know each other's gender, since that matches the real world. However, none of our claims below would change if we assumed that leaders did not know their opponent's gender.

<sup>45</sup>Fey (2008) finds these numerical solutions in a game where costs of effort are private. Corchón (2007) shows that, in the complete information game, changes to costs and changes to valuations of the prize are isomorphic.

<sup>46</sup>These exact quantities can be solved for, integrating over the NMVW and CA distributions.

### A.1.5 Conflict Initiation Decision

To decide whether to initiate a conflict, a leader again considers whether the expected utility of initiating a conflict is greater than or equal to zero. Note that here, the probability of winning an inter-group conflict is *endogenous*; it is determined by how much effort each leader exerts.<sup>47</sup>

Leader  $i$  chooses to initiate against  $j$  if:

$$\begin{aligned}
 p_i(e_i^*, \bar{e}_{MM}^*)(w + v_i) - k - c_i - e_i^* &\geq 0 && \text{Male leader (i) vs. male opponent (j)} \\
 p_i(e_i^*, \bar{e}_{MF}^*)(w + v_i) - k - c_i - e_i^* &\geq 0 && \text{Male leader (i) vs. female opponent (j)} \\
 p_i(e_i^*, \bar{e}_{FM}^*)(w + v_i) - k - c_i - e_i^* &\geq 0 && \text{Female leader (i) vs. male opponent (j)} \\
 p_i(e_i^*, \bar{e}_{FF}^*)(w + v_i) - k - c_i - e_i^* &\geq 0 && \text{Female leader (i) vs. female opponent (j)}
 \end{aligned}$$

These equations begin to demonstrate the two, competing effects of gender on conflict initiation. The most direct effect is from the  $c_i$  term. As in the decision to run for election, if women draw higher  $c_i$  terms, on average, then they will tend to initiate fewer conflicts.

However, the fact that only higher-NMVW women are likely to run for election – and that this effect is more stark for women than men – also means that female leaders tend to have higher NMVWs than males. This, in turn, means that (1) they are likely to have a higher  $v_i$  and (2) they tend to have higher probabilities of winning inter-group conflicts, since  $e_i^*$  is increasing in  $v_i$ .

Which effect dominates at the aggregate level depends on the parameters in the game, as we show below. Under some conditions, women initiate more conflicts, and under other conditions, they initiate fewer conflicts.

## A.2 Analysis

This is a contest game with incomplete information. Unfortunately, such games generally do *not* have closed form solutions for effort levels in equilibrium.<sup>48</sup> We therefore use simulations to show the effect of changing parameter values. In what follows, we set the number of players to  $N = 100,000$ , half men and women. We set the initial distribution of costs to entering contests as  $c \sim U[10, 30]$  for men. We set the distribution of NMVWs as  $v \sim U[0, 100]$ . In the following sections, we vary particular parameters and note the value of any others, where appropriate.

### A.2.1 Effect of Z: Difference in Cost Distributions

We begin by illustrating the main argument from our research note. This result hinges on the size of  $z$ , which captures how much more costly it is for women (on average) to enter a competition than men. Recall that the cost of candidacy (and conflict initiation) for men is given by  $c \sim U[\underline{c}_M, \bar{c}_M]$ , and  $\underline{c}_M > 0$ . For women,  $c \sim U[\underline{c}_M + z, \bar{c}_M + z]$ , where  $z > 0$ . As  $z$  increases, women draw their costs from a distribution of higher costs than men. As  $z$  goes to zero, this is like the world, at both

<sup>47</sup>This is in interesting contrast with models based on crisis bargaining, like Reiter and Wolford (2022), where the probability that one side wins is exogenously determined by the two sides' relative, *ex ante* capabilities.

<sup>48</sup>Fey (2008).

the domestic electoral level and the international level, becoming more “equal,” with women and men facing identical constraints.

The research note argues that the differential in candidacy costs - captured by  $z$  - has three observational implications. We illustrate each of these in Figure A.1. First, Panel (a) shows that the rate at which women run for office decreases as  $z$  increases. When women have to pay a high cost to run for office, fewer draw a high enough NMVW value to justify running for office.  $z$  has no effect on the rates at which men run, since it isn't a factor in their cost-benefit calculations. Second, as Panel (b) shows, the size of  $z$  also affects the average NMVW of the women who do end up running. As  $z$  increases, it takes women a higher NMVW to justify running. The NMVWs of women who run are increasingly greater than men who choose to run.

Finally, Panel (c) illustrates the consequence for interstate conflict that we highlight in the research note: the higher NMVWs resulting from a larger  $z$  also lead to a greater effort level in any ensuing inter-group conflict. To describe effort levels, we have to first specify “effort against whom,” since optimal effort levels differ depending on whether you are facing a female or male opponent. In panel (c), we show optimal effort for men and women when facing a male opponent. The figure is very similar for leaders facing a female opponent, just with the two lines shifted vertically. We also have to specify “effort by whom,” since some leaders do and do not want to initiate conflicts.

In panel (c), we pooled all male and female leaders, regardless of whether they would have chosen to initiate a conflict. We show effort levels in this way first, because this matches the laboratory protocol, where every leader is automatically entered into a conflict.<sup>49</sup> In section 2.3, we show all possible dyads, of initiators and non-initiators, against male and female opponents. Female leaders still fight harder.

## A.2.2 Alternative Differences Across Genders

As discussed in the main manuscript, there are reasons to believe there is a gender gap in the costs of running or initiating conflict. However, gender differences may also exist in the probability of winning an election and/or the common benefits to winning an election or conflict. In this section, we demonstrate that the effects of such differences are qualitatively the same as the effects of changing the costs to entering a contest, as in the preceding section.

First, we allow the probability of winning an election - real or perceived - to differ by gender. In the above model,  $p$  was common to all players. Here, we scale  $p$  by some factor between 0 and 1. In other words, we decrease the probability of winning an election used by women in their cost-benefit calculation, without changing anything about the ensuing inter-group conflict. When  $p$  is small women believe they are much less likely to be elected than men. As  $p$  goes to 1, this difference shrinks.<sup>50</sup>

The left hand column of Figure A.2 shows the effects on the running rates of individuals by gender, the average NMVW of leaders by gender, and the average effort in an inter-group conflict by gender. When there is a greater gender gap - i.e. when the  $p$  scaler is smaller - the difference between men and women with respect to all three outcomes is greater.

---

<sup>49</sup>In our protocol, participants didn't know their opponent's gender. Though note that this figure would again be qualitatively the same if we removed the “against whom” aspect as well.

<sup>50</sup>We set  $z = 5$  for the simulations in this section.

Second, we allow the value of winning an election - again, real or perceived - to differ by gender. This allows for the possibility that women do not value the “prize” of winning office as highly as men. Again, we scale the relevant parameter,  $w$ , by a factor between 0 and 1. Recall,  $w$  represents the value of winning an election that is common to all players, aside from their individually drawn NMVW value. The right hand column of Figure A.2 shows how shrinking this benefit for women impacts candidacy rates, average NMVW and effort in intergroup contests. As above, a greater gender gap - i.e. a smaller  $w$  scaler - results in a greater difference between men and women.

In short, these alternative gender differences all have the same impact as increasing the cost differential between genders. Anything that makes running for office less attractive for women means that candidates must have a higher NMVW to justify running. So fewer women run. Among those that do, they have higher NMVWs and ensuing effort levels in a conflict, compared to men that run.

### A.2.3 Adding Endogenous Conflict Initiation

The preceding two sections showed how effort levels varied by for male and female leaders, but it did not break those effort levels down by initiators and non-initiators. Here, we take full account of the effect of endogenous conflict initiation to show how our main claim - effort levels of female leaders are higher - obtains.

Figure A.3 shows the effort levels for all possible leaders in all situations. The top two panes show the effort levels of a leader who chose to initiate a conflict against a male opponent (left) and a female opponent (right). The bottom two panes show the same thing for a leader who would not have chosen to initiate, but who nonetheless found themselves in a conflict with an opposing leader who initiated.

In all cases, a female leader fights harder than a male leader - on average - and this difference is increasing in  $z$ . Therefore, regardless of who initiates the conflict and regardless of the gender of the opponent, women leaders’ effort levels are higher.

Note that the red lines - for male leaders - are flat on the left hand side and upward sloping on the right hand side. This is because the left hand side considers male opponents, so  $z$  does not affect the attributes of the pool of male leaders. The lines are upward sloping on the right hand side because, as  $z$  increases, female opponents have higher NMVWs on average, so males increase their effort levels to account for this.

### A.2.4 Initiation Rates and Relative Sexism at the International vs. Domestic Level

We now directly analyze initiation rates by gender. Our goal is to show that initiation rates can be higher for men or for women, depending on assumptions about the relative degree of gendered cost differentials at the domestic versus international level. Put simply, if the degree to which men and women face different costs to entering a contest is the same at the domestic/electoral level as it is at the international/conflict level, then conflict initiation rates are lower for women (even though their effort in conflicts is still higher, as shown above). However, if the gendered cost differentials are sufficiently large at the domestic level, relative to the international level, then women may initiate conflicts at higher rates than men.

Recall, that there are two competing effects of gender on conflict initiation rates. There is the direct effect: if women draw higher conflict initiation costs, on average, then they are less likely to start conflicts. But there is also the selection effect: if only the highest NMVW women enter elections, then they know that they will also fight harder in an interstate conflict and be more likely to win, which makes them more likely to start conflicts. If the selection effect is strong enough because the electoral environment has bigger gender cost differentials, then it can dominate the direct effect.

We extend the model above to introduce the possibility that the gender difference in candidacy costs at the electoral stage may be bigger *or* smaller than the gender difference in the costs of conflict initiation. We operationalize this by adding or subtracting a fixed amount from conflict initiation costs at the international level. Recall that the distributions of cost draws were  $c \sim U[\underline{c}_M, \bar{c}_M]$  for men, and  $c \sim U[\underline{c}_M + z, \bar{c}_M + z]$ , where  $z > 0$ , for women. And we previously assumed that this cost draw applied to the election and a possible interstate conflict.

Now, assume that costs at the interstate conflict level, a female player's costs are  $c_i + b$ .  $b$  can be positive or negative. If  $b > 0$ , then the degree to which women face higher contest costs is even higher at the international than the domestic level. If  $b < 0$ , then costs are higher at the domestic level.

For example, suppose costs were partially driven by psychological competition aversion and partially driven by party leader gate-keeping. If party leaders affirmatively sought out and encouraged female candidates, then this could offset costs at the domestic level but not the international level ( $b > 0$ ). Conversely, if party leaders were stricter gate-keepers against women candidates, but the female leader didn't need party backing as much for her foreign policy decisions, then costs could be lower at the international level ( $b < 0$ ).

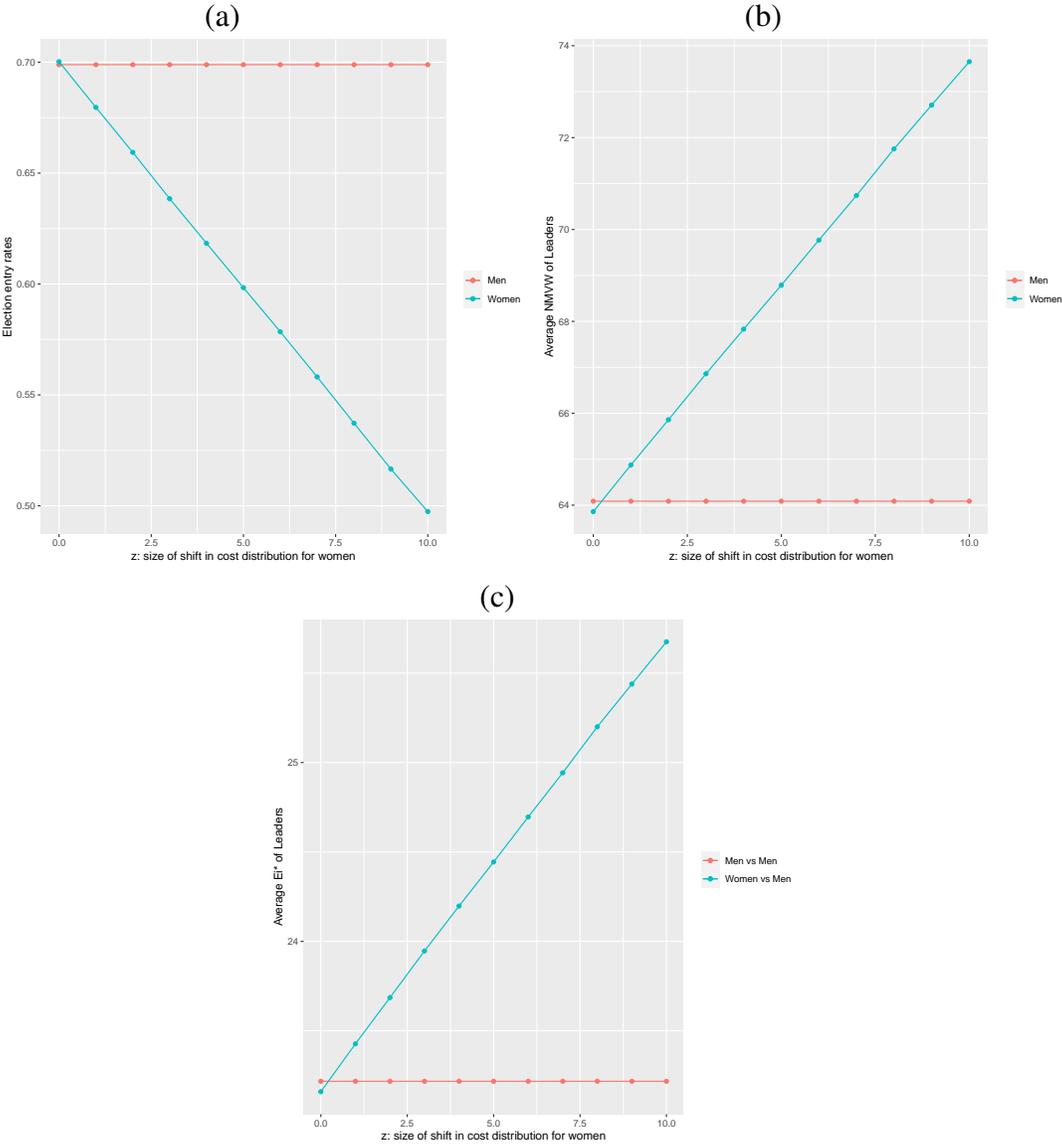
Figure A.4 shows how conflict initiation rates by gender vary as  $b$  changes. On the right hand side of the figure, a female leader's costs to starting an international conflict are even greater than her costs to entering an election. There, the direct effect of conflict costs dominate. Though female leaders have higher NMVWs, they are even more conflict averse at the international level than men.

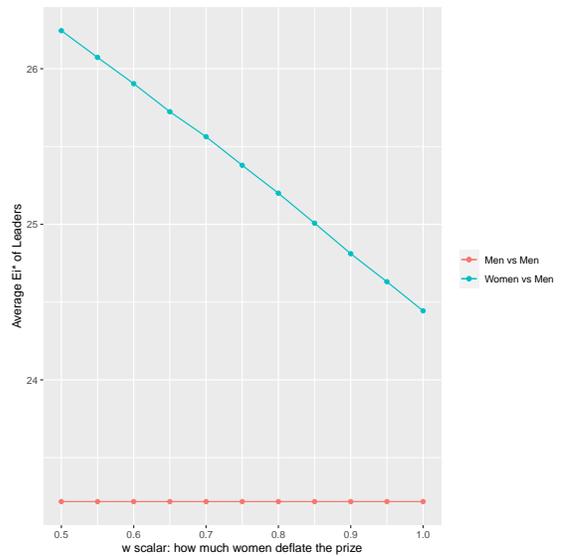
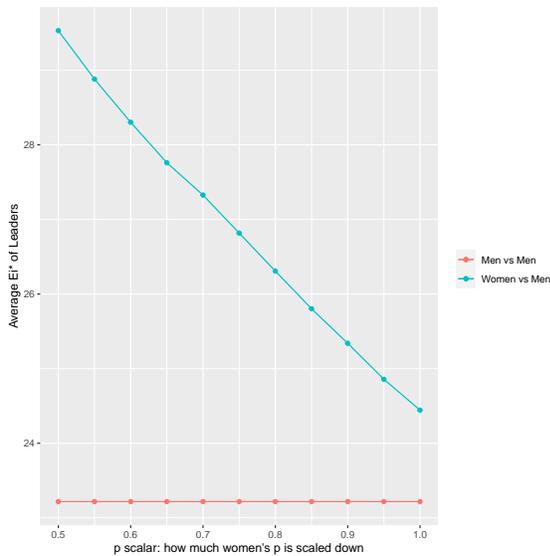
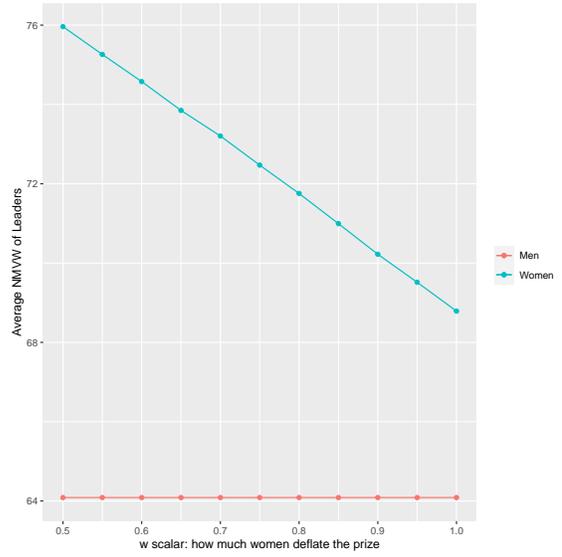
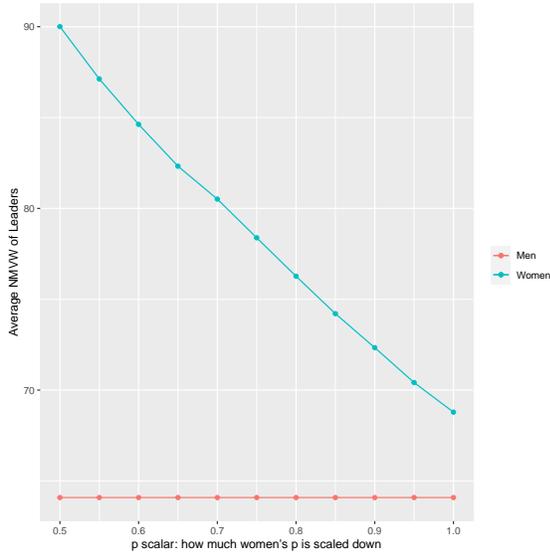
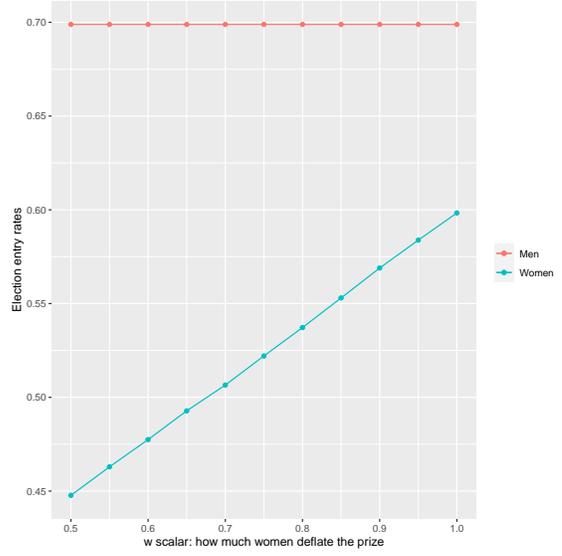
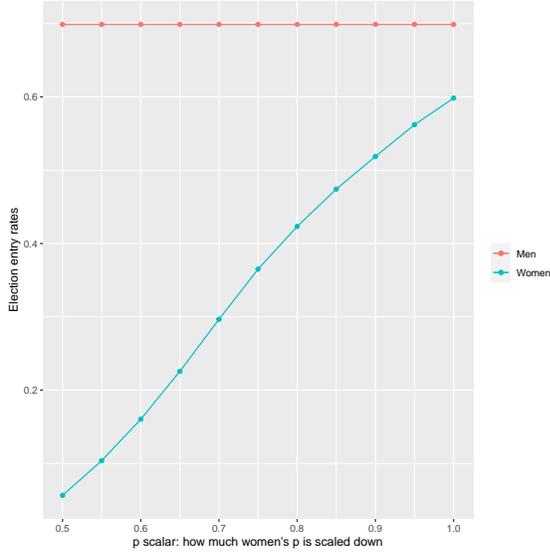
On the left side of this figure, where  $b$  is negative, women face smaller conflict initiation costs, compared to their election entry costs. Here, the selection effect dominates. Female leaders have already overcome the steeper costs of entering an election, so when they become leaders, this constraint is removed. However, they still have their high NMVW draw, which propelled them to run in the first place. Figure A.4 showed initiation rates against male opponents. The figure is similar for initiation rates against female opponents.

Note that at  $b = 0$ , where there are no differences between the two levels, women initiate at lower rates than men. Though, also note that the crossing point in this figure is at  $b = -2$ , which is *less than*  $z$ . Here, we set  $z = 5$ . So female leaders are initiating at a higher rate than males, even when there is still a cost differential at both levels. The simulation in Figure A.4 is not one in which women have been subsidized with lower costs to interstate conflict initiation.

Figure A.5 shows the effort levels of leaders against male opponents, for initiators and non-initiators. Again, for all values of  $b$ , female leaders choose higher effort levels than males. Note that in the left hand figure, for initiators, female and male leaders who choose to initiate against males have the same effort levels when  $b = 5$ . This is as expected - when  $b = z = 5$ , the gendered cost differential at the electoral level has been offset at the international level. Among initiators, women and men have the same NMVW, in expectation.

Figure A.1: Effect of gendered cost differences on running rates, NMVW, and effort in conflicts

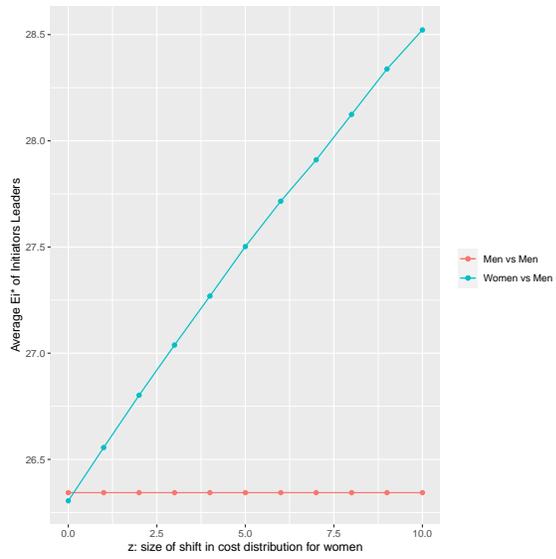




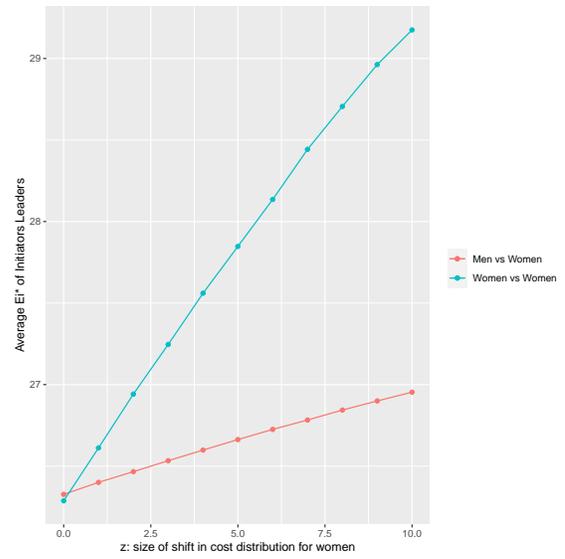
Pr(win) Election

Common Benefit

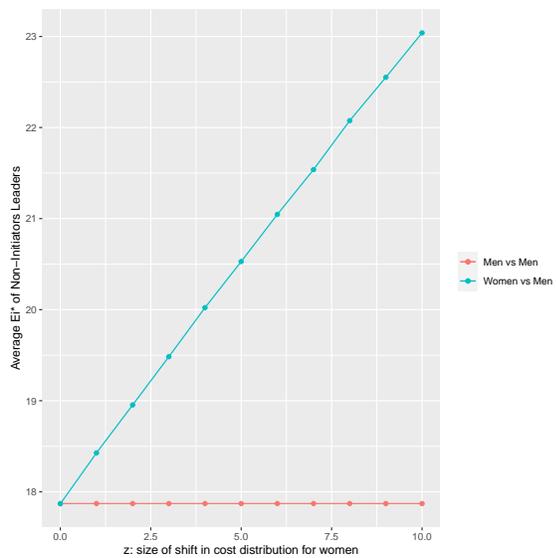
Figure A.2: Effect of changing the probability of winning, common benefit to winning



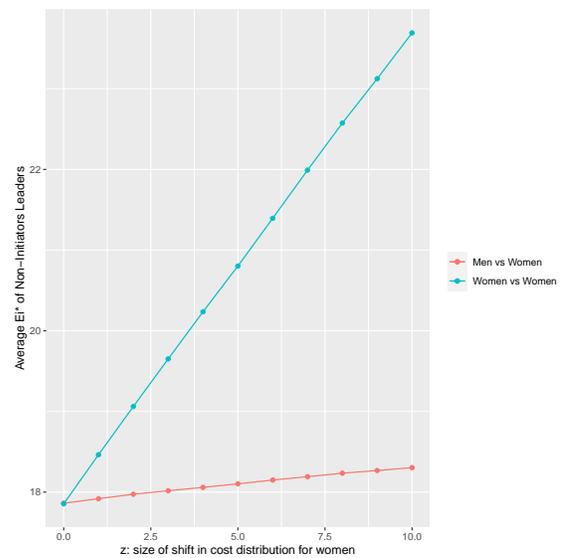
(a) Initiators, vs. Male Opponents



(b) Initiators, vs. Female Opponents



(c) Non-Initiators, vs. Male Opponents



(d) Non-Initiators, vs. Female Opponents

Figure A.3: Effect of gendered cost differences on effort in conflicts, varying opponent and initiator status

Figure A.4: Effect of varying  $b$  on conflict initiation rates, male opponent

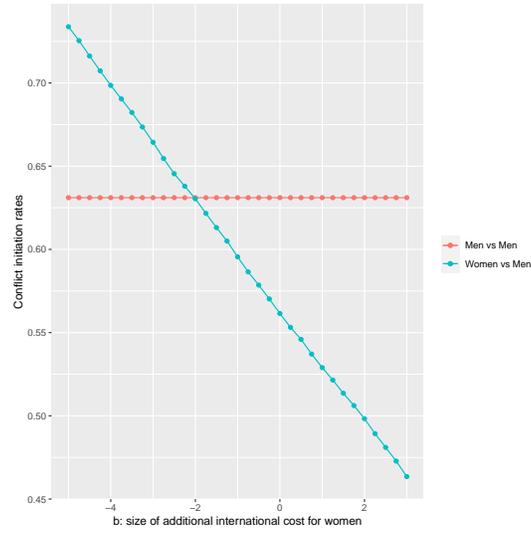
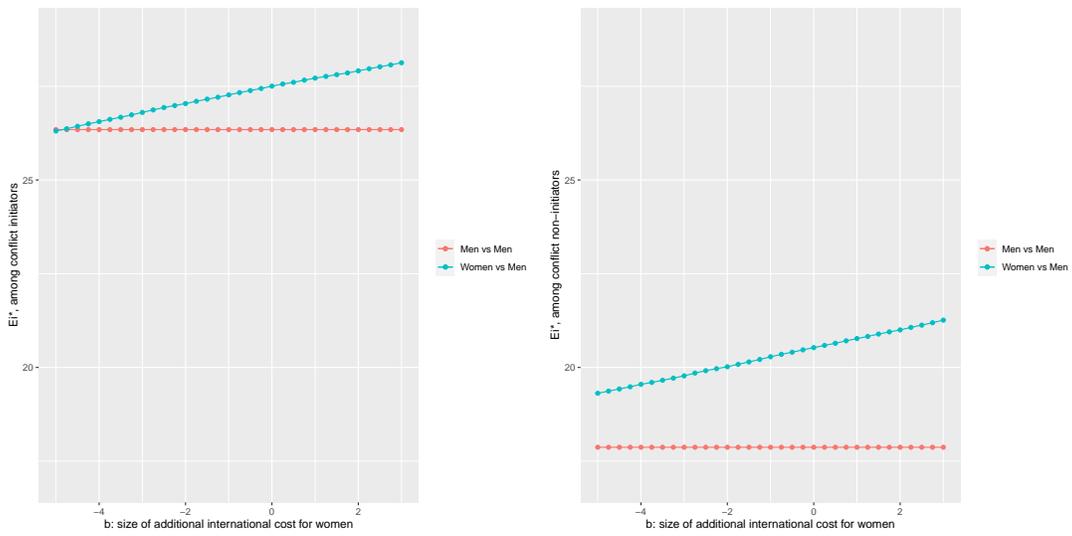


Figure A.5: Effect of varying  $b$  on effort, male opponent



### A.2.5 Effect of Relative Domestic vs. International Stakes

In the previous section, we varied the degree to which the domestic and international levels had gendered cost differentials. The two levels obviously differ in other important ways. The main way is that the “stakes” at the international level could be bigger or smaller than the stakes of an election. Here, we show that varying the stakes of the two levels (a) does not change our main results about effort and (b) is insufficient to change whether male or female leaders initiate conflict at higher rates.

By “stakes,” we mean the relative magnitude of potential gains/losses and relative marginal costs to effort at the domestic versus international level. For example, the costs of losing an election (or conversely, the private benefits to winning an election) are meaningful. But they are much smaller than the costs of losing an existential interstate conflict. Switching to the marginal cost of effort, it is unpleasant to spend money on a campaign, but it is especially painful to spill blood in a war.

Here, we show how initiation rates and effort change as we vary the relative stakes of an election versus a conflict. To operationalize this, recall that  $w$  described the part of winning a contest that was beneficial to everyone, aside from an individual’s NMVW. For an election, this is the pride or financial gain from winning office that every winner enjoys. For an inter-group conflict, this is the prestige or legacy boost or material gain that every victorious leader enjoys. We previously assumed that  $w$  took on the same value in the electoral stage as in the inter-group conflict stage.

Here, we scale  $w$  by some positive factor for the inter-group conflict. This is akin to saying “how much bigger or smaller are the common stakes of an inter-group conflict, compared to the stakes of an election?” When this scalar is greater than 1, the stakes of the interstate conflict are greater than the electoral stakes. When the scalar is less than 1, the election is higher stakes.

Figure A.6 shows that varying stakes still does not overturn our prediction that female leaders will, on average, have higher effort levels in conflict than male leaders. At every point as we vary the relative stakes, female leaders still fight harder. This is because, scaling the conflict prize upwards for everyone increases everyone’s effort level, but it still does not change the fact that gendered selection in the electoral stage means that female leaders still tend to have higher NMVWs than male leaders. So their total “prize” - individual specific value to winning plus the common prize - is greater than the males, on average, resulting in higher effort levels.

Figure A.7 shows how varying the stakes affects conflict initiation rates against men.<sup>51</sup> Men begin with higher initiation rates than women, but as the relative stakes of the inter-group conflict increase, the initiation rates converge. This makes sense. If the inter-group conflict prize is high enough, both genders, regardless of their individual-specific NMVWs will choose to initiate conflict, shrinking the gendered initiation gap. Note that here, we have re-set  $b = 0$ , so that there is no difference in the degree to which costs vary by gender at the domestic versus international levels. If we chose  $b = -3$ , for example, this figure would be similar, but female leaders would start at higher initiation rates than male leaders.

---

<sup>51</sup>The figure is similar for initiation rates versus women.

Figure A.6: Effect of scaling common value to winning on effort in conflict

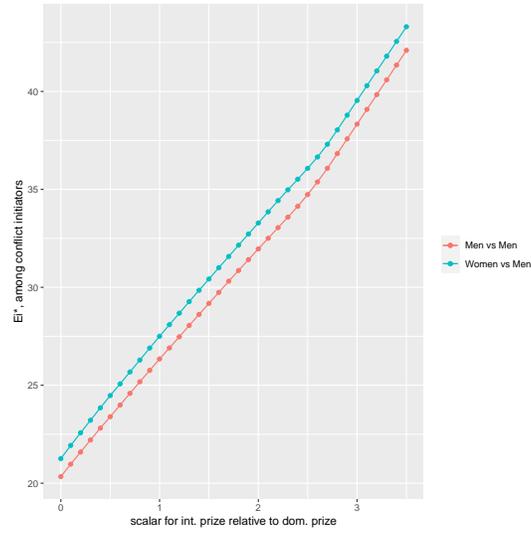
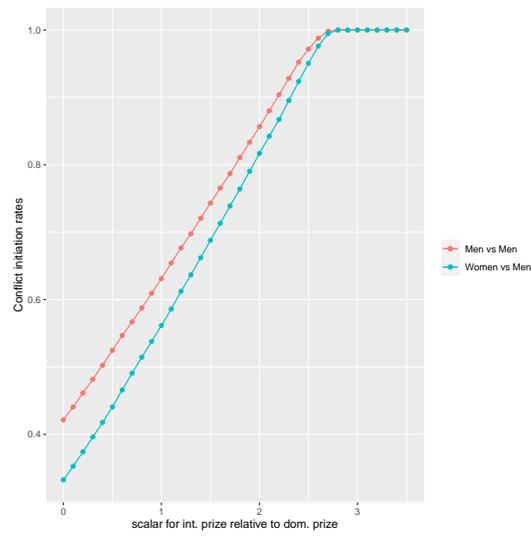


Figure A.7: Effect of scaling common value to winning on initiation rates



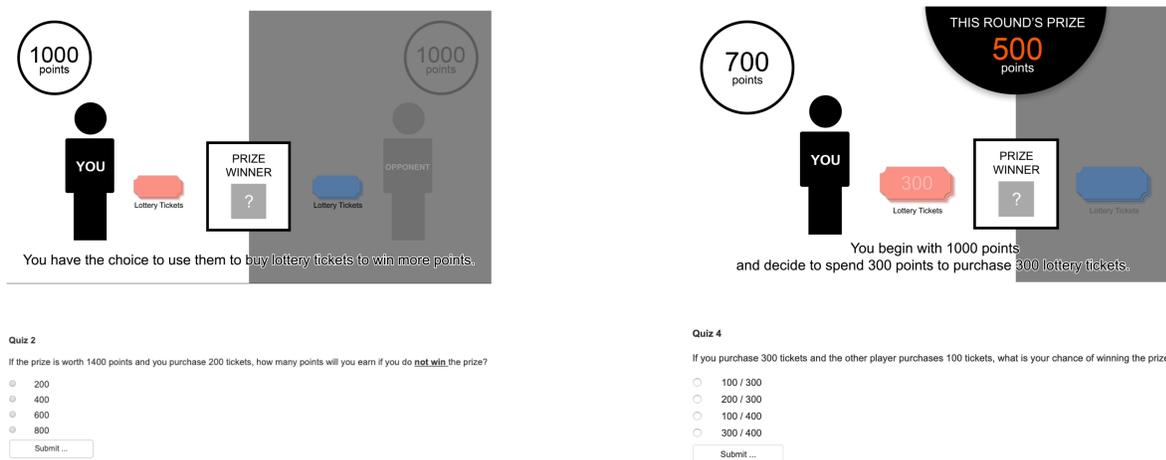
## B Appendix: Full Protocol Description, Recruitment, Demographics

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

### B.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.<sup>52</sup> Participants then answered four quiz questions to ensure that they understood the mapping between their choices and payoffs. Figure B.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 B.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 B.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

<sup>52</sup>The video is available here: <https://www.youtube.com/watch?v=3ywZvA0CLy8>.

earnings for that round.<sup>53</sup>

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

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

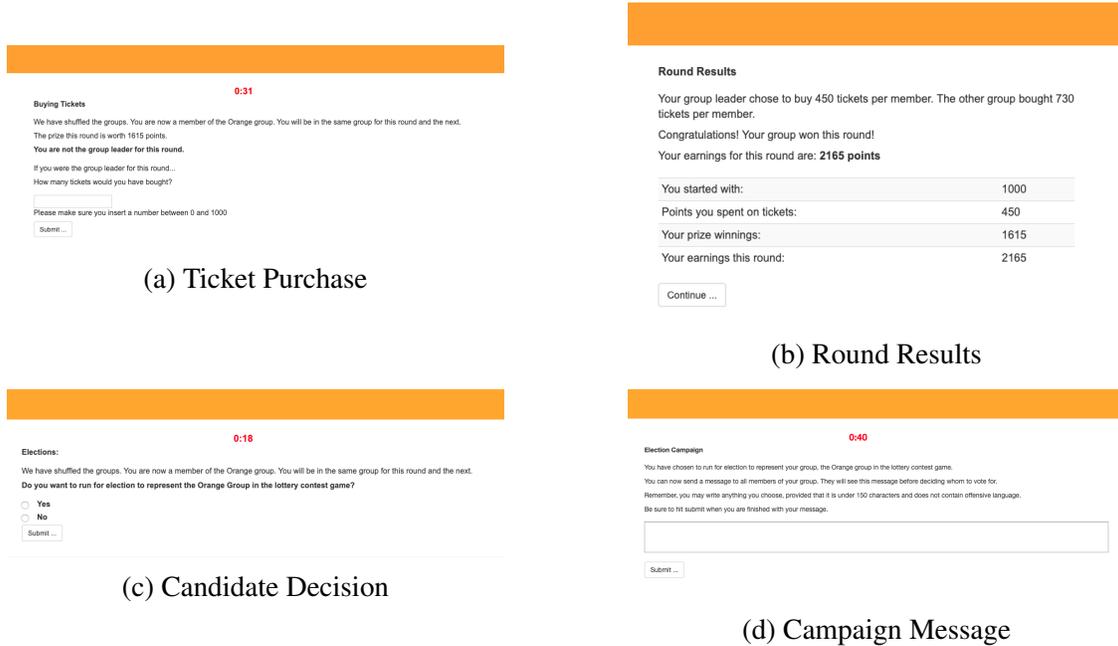


Table B.1 shows the prize values for each round and each part of the game. We chose these values to make sure that the prize value was sufficiently enticing to get people to bid, based on test runs. We picked values that weren’t exactly round numbers, like 2000, in order to decrease the power of focal 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.<sup>54</sup>

<sup>53</sup>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.

<sup>54</sup>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 B.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

## B.2 Compensation and Timing

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.

Participants received \$18, approximately, on average. This is a slight overestimate of their pay, because we over-recruited for each session to make sure that we had enough participants. If a participant was dismissed without playing the game, we still paid them the \$5 show up fee. The game itself, including instructions, took approximately 35-45 minutes per session.

## B.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 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 finish their timed tasks before they are placed into pairs or groups.

Recruitment for higher-paying tasks like ours was tricky, because most Turkers use automated scripts that claim *any* tasks (HITs - Human Interface Tasks, in MTurk terminology) that meet the Turker's specified criteria. HITs that are claimed by scripts *do not* have to be completed

immediately, so we couldn't simply advertise a high-paying HIT at a particular time. We instead had to advertise a \$0.01 HIT whose description said that it was a recruitment HIT for a game to be played at a particular, pre-specified time. Turkers who completed the recruitment HIT were then messaged an individual link that took them to the game, at their designated time. Since the automated scripts didn't immediately grab the \$0.01 HITs, this meant that Turkers weren't using their scripts to grab one of our HITs and squat on it. And we could actually coordinate to get people to the right place at the right time. A Turker could only play the game once, regardless of sessions, and MTurk lets the researcher specify that the same Turker cannot complete more than one HIT in a particular "batch."

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.<sup>55</sup> For those who did show up 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.

Few participants dropped out. Across all sessions, seven participants dropped out before the second round of the ICG and two more dropped out in at a later point.

## **B.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 B.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.<sup>56</sup> 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.

---

<sup>55</sup>For those who showed up late, we still paid the \$5 show up fee even though we dismissed them from the study.

<sup>56</sup>CCES and our MTurk sample measure household income, and Anderson et al. (2013) ask for the participants' parents' income.

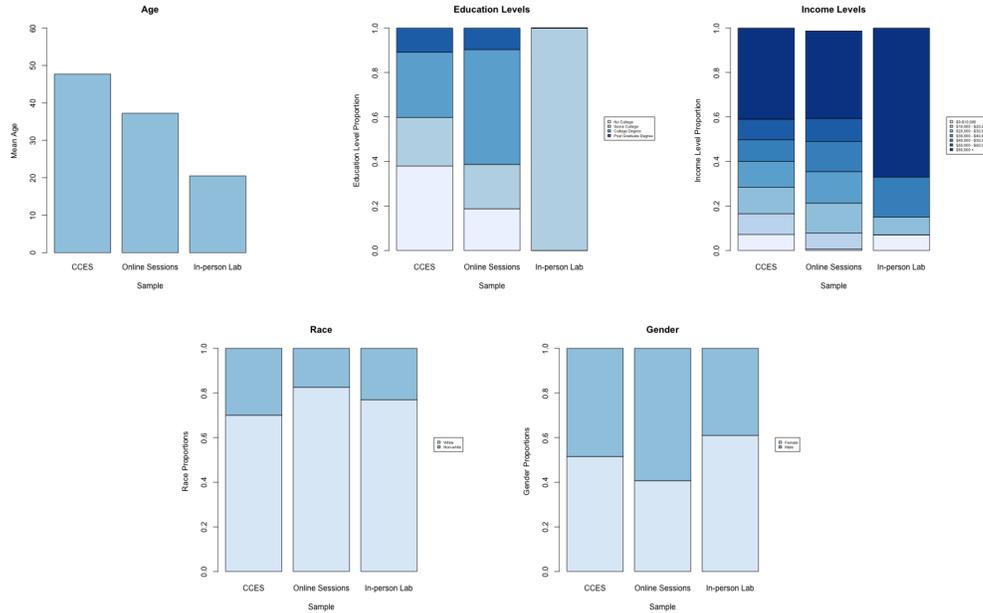


Figure B.3: Comparison of Sample Characteristics

## C 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.

### C.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.

## **C.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.

### C.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 C.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 C.1: Binomial Distribution for Random Leader Selection

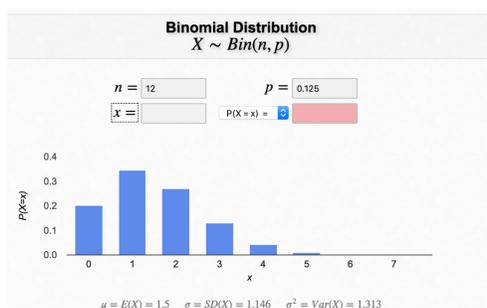


Figure Source: Bognar, Matt. Department of Statistics and Actuarial Science, University of Iowa. <https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>.

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.

## D Appendix: Differences in Leader Behavior by Gender

In Table D.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 D.1: Effect of Gender on Tickets Bought

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

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table D.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*** (14.652)	-28.923** (12.813)	-35.463*** (12.813)	-27.646** (13.057)	-34.186*** (13.056)
DSG	60.762** (24.677)	57.221*** (21.673)	57.228*** (21.567)	50.503** (23.932)	50.522** (23.815)
Female X DSG	154.021*** (47.564)	126.011*** (41.238)	112.302*** (42.730)	132.404*** (42.752)	118.703*** (44.209)
Constant	523.295*** (16.437)	124.143*** (15.542)	124.963*** (15.516)	128.091*** (16.028)	128.992*** (16.004)
Observations	1,700	2,040	2,021	1,960	1,941

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## E Appendix: Robustness of Effect of NMVW by Gender

### E.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 E.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 E.1: Effect of NMVW on Leadership and Candidacy, by Gender

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

## E.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.

## F Appendix: Campaign Messages Coding and Analysis

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 first coded each message as being one of the ten message types given below, with an actual example for each. 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!”
- 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 F.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 F.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%
<i>Women</i>										
Primary Cat.	17 8%	24 11%	10 5%	46 21%	44 20%	22 10%	24 11%	9 4%	24 11%	0 0%
Individuals	11	12	5	17	24	13	8	3	13	0
Ave. NMVW	402	942	-156	1024	360	794	1269	210	876	NA
<i>Men</i>										
Primary Cat.	17 4%	83 20%	23 6%	27 6%	118 28%	32 8%	13 3%	4 1%	72 17%	29 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).

Table F.2 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.<sup>57</sup>

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 F.3 replicates Table F.2 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.

<sup>57</sup>See Appendix F for summary data about message types and for similar results using coding for whether the message has any content that fits a particular category.

Table F.2: Effect of Gender and NMVW on Winning, Controlling for Campaign Messages

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

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

Table F.3: Effect of Gender and NMVW on Winning, Controlling for Campaign Messages (Any content coding)

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

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

## G 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.

### G.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.<sup>58</sup> 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 G.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 G.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 G.2.

Table G.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

---

<sup>58</sup>Sahm (2017).

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

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Av. Perc change in tickets	51.41	37.26	7.65	17.67
<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Av. Perc change in tickets	46.96	35.53	11.99	19.95
<i>(III) Winning vs. Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
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 G.2: Differences in ICG behavior, Risk Aversion (1605 and 1225 rounds), by Gender

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Av. Perc change in tickets	0.25	0.22	0.05	0.11
<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Av. Perc change in tickets	0.22	0.23	0.06	0.13
<i>(III) Winning vs. Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
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.

when 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).

## **G.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).

## **G.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.)

## **G.4 Confidence**

### **G.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 G.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 G.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 G.3: Differences Between Candidates and Not Candidates, by Gender

	<i>Dependent variable:</i>		
	Candidate		
	(1)	(2)	(3)
Female	−0.817*** (0.139)	−0.849*** (0.137)	−0.892*** (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)	0.008 (0.008)	0.007 (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*** (0.014)	0.081*** (0.014)	0.085*** (0.014)
Constant	−0.381*** (0.079)	−0.366*** (0.079)	−0.309*** (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 G.4: Differences in ICG payoff, by Gender

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Ave. ICG Payoff	1401	1308	1377	1310
<i>difference</i>		+93		+67

<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Ave. ICG Payoff	1346	1306	1328	1311
<i>difference</i>		+40		+17

<i>(III) Winning vs. Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
Ave. ICG Payoff	1388	1313	1369	1310
<i>difference</i>		+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 G.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 G.5: Effect of ICG Payoff on Candidacy, by Gender

	<i>Dependent variable:</i>		
	Candidate		
	(1)	(2)	(3)
Female	-0.013 (0.489)	-2.690*** (0.656)	-1.126** (0.527)
Ave. ICG Payoff	0.001** (0.0002)	0.001** (0.0002)	0.001** (0.0002)
Ave. ICG Tickets		0.022 (0.031)	
Nash NMVW			0.006 (0.008)
Female × Ave. ICG Payoff	-0.0004 (0.0004)	0.0001 (0.0004)	0.0001 (0.0004)
Female × Ave. ICG Tickets		0.364*** (0.058)	
Female × Nash NMVW			0.091*** (0.014)
Constant	-1.039*** (0.320)	-1.204*** (0.395)	-1.108*** (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 G.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 G.6: Effect of ICG Payoffs on Electoral Success, by Gender

	<i>Dependent variable:</i>		
	Winning Candidate		
	(1)	(2)	(3)
Female	-0.230 (0.963)	-2.145 (1.319)	-1.091 (1.066)
Ave. ICG Payoff	0.001** (0.0004)	0.001*** (0.0004)	0.001*** (0.0004)
Ave. ICG Tickets		0.036 (0.052)	
Nash NMVW			0.009 (0.013)
Female × Ave. ICG Payoff	-0.0001 (0.001)	0.0001 (0.001)	0.0001 (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 G.7: Effect of Lagged Payoffs on Ticket Purchases, by Gender

	<i>Dependent variable:</i>	
	(1)	(2)
Female	−30.088 (29.004)	−39.383* (22.494)
Payoff (lagged)	−0.051*** (0.011)	−0.029*** (0.009)
Female x Payoff (lagged)	0.002 (0.018)	0.008 (0.014)
Observations	1,650	1,650
R <sup>2</sup>	0.022	0.414
Prize FE	No	Yes
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

#### G.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. Pruyzers 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 G.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  $t - 1$ .<sup>59</sup> 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

<sup>59</sup>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 G.8: Effect of Electoral Wins/Losses on Candidacy/Winning, by Gender

	Candidate	Electoral Success
Female	-0.554*** (0.165)	-0.237 (0.350)
Winning Candidate (lagged)	2.205*** (0.208)	0.821*** (0.284)
Losing Candidate (lagged)	1.598*** (0.161)	-0.025 (0.267)
Female x Winner (lagged)	0.321 (0.375)	-0.394 (0.528)
Female x Loser (Lagged)	0.599** (0.265)	-0.102 (0.464)
Observations	1,636	571
Log Likelihood	-898.695	-344.372
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

## G.5 Learning Over Time

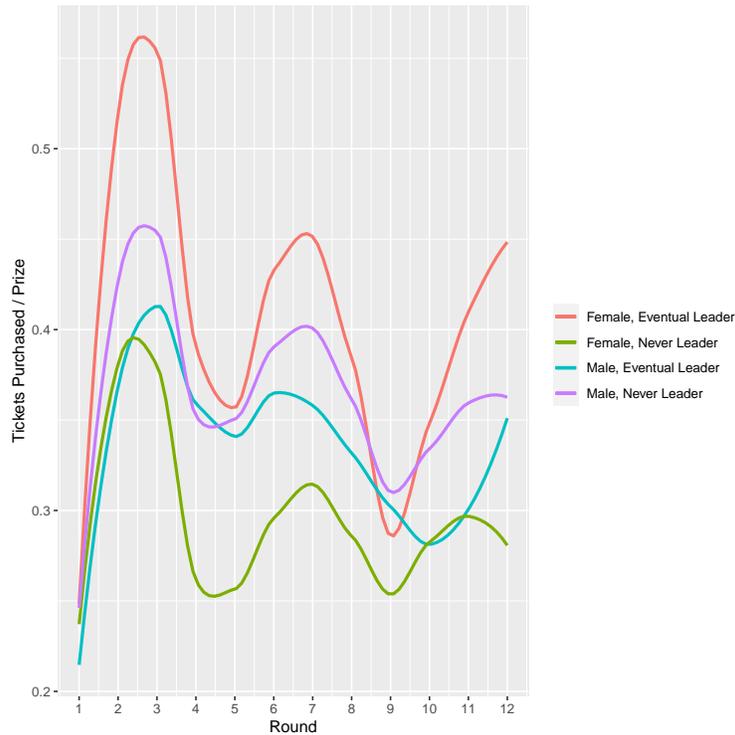
We thank a reviewer for pointing out that participants have more information available to them than just their payoffs. They also observe the decisions of the other players. Here, we investigate whether men and women appear to learn differently over time after observing others' play, and whether this also differs between those who are eventually an elected leader and those who are not elected leaders.

Figure G.1 shows data from the ICG. In each round, for each participant, we calculated the number of tickets they purchase as a percentage of the prize. We then divided the participants into four categories - men/women, eventual leaders/non-eventual leaders. In each round, there are multiple data points for each category of participant. Figure G.1 shows smoothed lines for each category across rounds. This gives a visual summary of how ticket purchases for each category of participant changed as rounds progressed.

First, the patterns match those highlighted in the main text. Eventually-elected women tend to have the highest ticket purchases across rounds, followed by both elected and never-elected men, and then women who were never elected. Second, there does not appear to be differential changes in behavior over the 12 rounds, according to group. All four groups slightly decrease their ticket purchases. It is not the case that one group of the other has distinct trends from the others.

We also looked for any statistical differences among the groups in learning over time. In Table G.9 we regressed the number of tickets purchased on a round counter, an indicator for female participants, and the interaction between the two. The first column is for eventual leaders, and the second column is for those who were never elected. There does not appear to be differential

Figure G.1: Learning over time in the ICG, by gender and eventual leadership



learning from observed play between groups. Columns 3 and 4 repeat the same process for those who eventually do and do not become candidates.

## G.6 Group Play

We thank a reviewer for also highlighting that the ICG and DSG differ in two ways. The latter has the presence of a group leader making a choice on behalf of her group *and* that leader is chosen via election. In other words, we have added groups and endogenous leader selection. We designed an additional part of our protocol to address this issue, which lets us show that the main effects we highlight are driven by the presence of elections.

In between the ICG and the DSG, we also put players into groups of the same size as the DSG. However, rather than elections, we *randomly* selected a person from each group to be their group's leader. We called this the RSG. Comparing data from the RSG and the DSG allows us to isolate the effect of democratic leader selection and distinguish it from any effects that stem solely from group player or the presence of leaders. We didn't emphasize this part of the protocol or data analysis in the main manuscript because of space constraints. Since, as explained momentarily, the results are largely the same, we left the main manuscript's emphasis on comparing the ICG and the DSG.

In what follows, we use data from the RSG to calculate our same measures of NMVW. In the main manuscript, we calculated these measures based on individual decisions in the ICG. Here, we do the same with their decisions from the RSG. Though, note one difference: when a player

Table G.9: Learning over time, by gender and eventual leadership/candidacy

	<i>Eventual Leaders</i>	<i>Never Leaders</i>	<i>Eventual Cand.</i>	<i>Nev. Cand.</i>
	boughtTickets			
	(1)	(2)	(3)	(4)
Female	53.274 (49.347)	-101.819** (48.890)	1.291 (39.902)	-120.420 (74.974)
Round number	2.605 (3.992)	3.759 (4.529)	2.957 (3.256)	4.096 (7.938)
Female X Round number	1.784 (6.705)	-1.954 (6.643)	1.946 (5.422)	-6.054 (10.187)
Constant	532.404*** (29.378)	553.450*** (33.331)	537.884*** (23.963)	566.249*** (58.419)
Observations	948	852	1,464	336

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

was *not* chosen as leader, we asked them how many tickets they *would have chosen* had they been leader. We called this their “hypothetical decision.” For this analysis, we calculated - for each individual - their NMVW measures based on how many tickets they bought when they were, in fact, randomly chosen as leader and when they were asked for their hypothetical decision.

All of the main arguments in the original manuscript obtain using RSG-based measures. Women, overall, tended to have lower NMVWs based on RSG choices, but those with higher NMVWs were more likely to select into candidacy and win elections. And they subsequently chose higher ticket purchases in the DSG. And these selection effects were much more stark for women than for men. The only difference in results when using RSG measures as opposed to ICG measures is that - using RSG measures, the selection effects for women are more heavily concentrated in the candidacy decision, as opposed to who gets elected once they’ve selected into candidacy. Using ICG measures, the candidacy selection effect was still stronger, but the electoral selection effect was also more clearly present.

Table G.10 replicates Table 1 from the main manuscript. Using RSG data to calculate NMVW and the average number of tickets purchased, we can see that the men who did and did not become leaders eventually differed only slightly. On the other hand, women who eventually became elected leaders purchased many more tickets in the RSG, compared to women who were not eventually elected and compared to all men. The only notable difference is that selection effects are most strongly concentrated in the candidacy decision, and less so in terms of winning versus losing candidates. Using ICG channels, the relative strength of the two selection effects was more balanced.

Table G.11 replicates Table 2 from the main manuscript. It shows similar results. Higher

NMVWs and ticket purchases in the RSG was associated with a higher probability of eventual election, much more so for women than for men. Table G.11 replicates Table 2, which shows these same claims in terms of estimated effects.

Table G.12 replicates Table 3, showing how greater NMVW in the RSG was associated with writing longer campaign messages, and that this effect was stronger for women than for men. Table G.13 replicates Table F.2. Here, the main difference between results using the ICG and RSG data is that we cannot reject the null that NMVWs had the same effect for men versus women, controlling for message length and type. This is likely the reason why the candidacy selection stage seems to matter more when we use ICG data.

Finally, Table G.14 replicates Table 4, using the RSG data to measure risk aversion. We can again rule out the possibility that variation in risk aversion explains our results. If anything, women who were eventually elected leaders were slightly more risk averse than women who were not eventually elected.

Table G.10: Differences in NMVWs, by Gender

---



---

(I) Leaders vs. Not Leaders

	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Ave. Nash NMVW	567	594	1045	613
<i>difference</i>		+27		+432
Ave. RSG Tickets	513	513	622	505
<i>difference</i>		0		+117

---



---

(II) Candidates vs. Not Candidates

	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Ave. Nash NMVW	517	647	934	520
<i>difference</i>		-130		+414
Ave. RSG Tickets	498	525	594	479
<i>difference</i>		-27		+115

---



---

(III) Winning vs Losing Candidates

	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
Ave. Nash NMVW	519	487	1021	893
<i>difference</i>		+33		+128
Ave. RSG Tickets	502	488	615	584
<i>difference</i>		+14		+31

---



---

Note: Part III excludes candidates who ran unopposed.

Table G.11: Effect of NMVW on Leadership and Candidacy, by Gender

	<i>Dependent variable:</i>					
	Is Leader		Is Candidate		Is Winner	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem.	-1.406*** (0.243)	-2.547*** (0.526)	-1.315*** (0.151)	-2.947*** (0.344)	-0.626** (0.291)	-0.773 (0.619)
RSG Nash NMVW	-0.005 (0.013)		-0.031*** (0.010)		0.006 (0.014)	
Fem. X RSG Nash NMVW	0.081*** (0.022)		0.117*** (0.016)		0.017 (0.026)	
Ave. RSG Tickets		0.003 (0.050)		-0.101*** (0.038)		0.040 (0.057)
Fem. X Ave RSG Tick.		0.318*** (0.087)		0.458*** (0.061)		0.050 (0.103)
Constant	-1.574*** (0.110)	-1.615*** (0.269)	-0.095 (0.084)	0.240 (0.203)	-0.728*** (0.126)	-0.893*** (0.301)
Observations	1,786	1,786	1,786	1,786	645	645

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table G.12: Length of Campaign Messages, by Effort and Gender, using RSG data

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

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

Table G.13: Effect of Gender and RSG NMVW on Winning, Controlling for Campaign Messages

	(1)	(2)	(3)
RSG NMVW	0.018 (0.015)	-0.004 (0.017)	0.005 (0.017)
Female	-0.482 (0.306)	-0.646** (0.321)	-0.568* (0.331)
Length	0.012*** (0.002)		0.008*** (0.003)
RSG NMVW x Female	0.005 (0.027)	0.026 (0.029)	0.020 (0.029)
Bid High		-0.377 (0.472)	-0.421 (0.467)
Bid Low		-1.113** (0.511)	-1.046** (0.505)
Humor		-0.016 (0.298)	-0.132 (0.304)
Critique		0.698** (0.322)	0.515 (0.329)
Skill		0.550** (0.259)	0.369 (0.268)
Speed		0.542 (0.504)	0.479 (0.509)
Strategy		1.211*** (0.386)	0.957** (0.394)
Team		0.307 (0.198)	0.122 (0.209)
Track Record		0.944*** (0.244)	0.825*** (0.248)
Constant	-1.626*** (0.223)	-1.469*** (0.276)	-1.806*** (0.303)
Observations	638	638	638

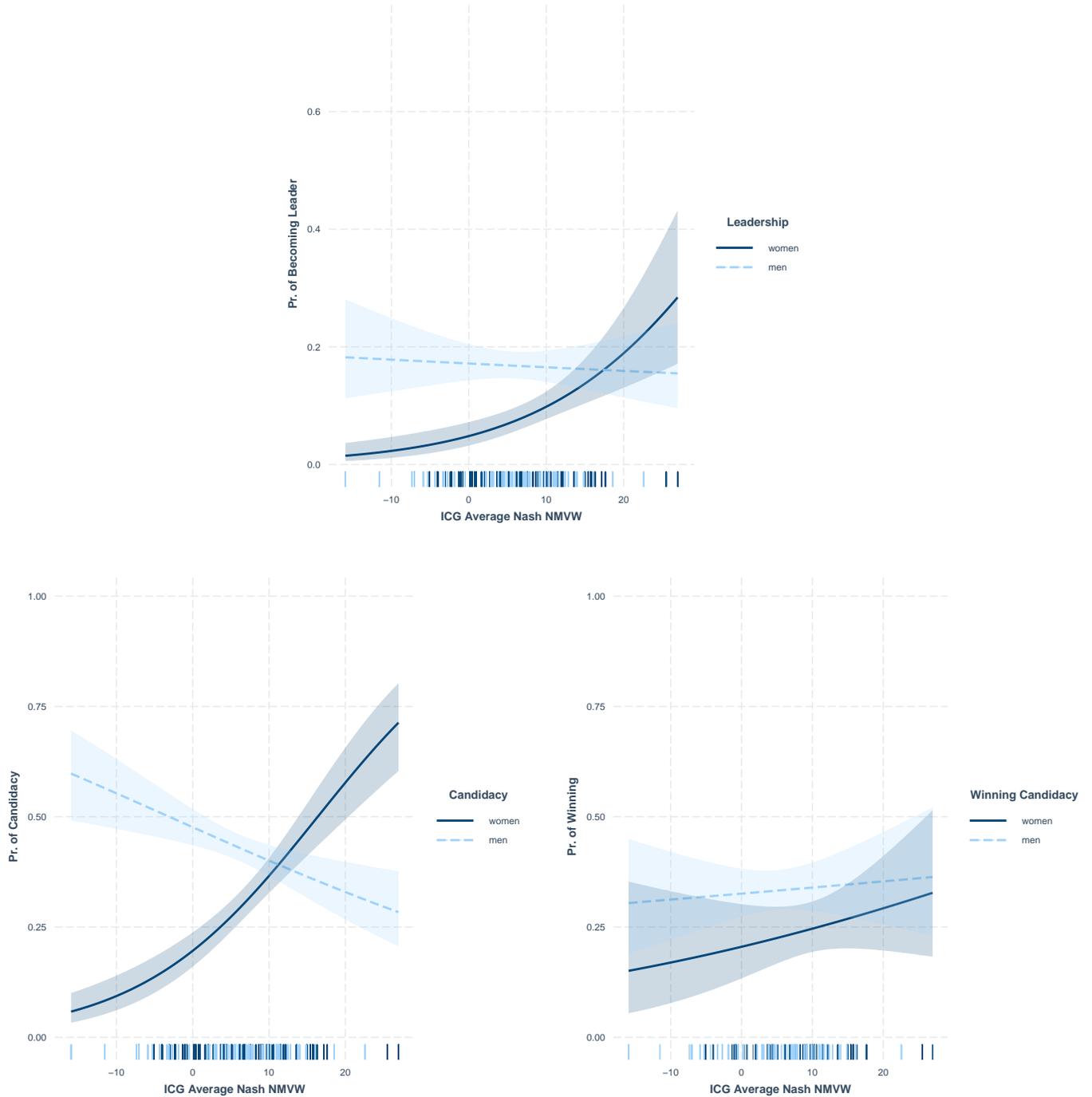
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table G.14: Differences in RSG behavior, Risk Aversion (LVR and 1225 rounds), by Gender

<i>(I) Leaders vs. Not Leaders</i>				
	<u>Men</u>		<u>Women</u>	
	<i>leader</i>	<i>not leader</i>	<i>leader</i>	<i>not leader</i>
Av. Perc change in tickets	24.51	14.82	10.71	13.54
<i>(II) Candidates vs. Not Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>candidate</i>	<i>not candidate</i>	<i>candidate</i>	<i>not candidate</i>
Av. Perc change in tickets	27.37	8.36	12.76	13.78
<i>(III) Winning vs. Losing Candidates</i>				
	<u>Men</u>		<u>Women</u>	
	<i>winner</i>	<i>loser</i>	<i>winner</i>	<i>loser</i>
Av. Perc change in tickets	23.20	29.15	11.12	13.51

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

Figure G.2: Predicted Effects of Nash NMVW, by Gender, using RSG data



Note: The plot shows predicted probability of being a leader or candidate in a particular round of the DSG. Predictions use estimates from Table G.11.

## 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.
- Ashworth, Scott, Christopher R Berry and Ethan Bueno de Mesquita. 2023. "Modeling Theories of Women's Underrepresentation in Elections." *American Journal of Political Science* .
- 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).
- 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.
- Corchón, Luis C. 2007. "The theory of contests: a survey." *Review of economic design* 11:69–100.
- Fey, Mark. 2008. "Rent-seeking contests with incomplete information." *Public Choice* 135:225–236.
- Fox, Richard and Jennifer Lawless. 2011. "Gendered Perceptions and Political Candidacies: A Central Barrier to Women's Equality in Electoral Politics." *American Journal of Political Science* 55(1):59–73.
- Hendriks, Achim. 2012. "SoPHIE-Software platform for human interaction experiments." *University of Osnabrueck, working paper* .
- 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.
- 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).
- Reiter, Dan and Scott Wolford. 2022. "Gender, sexism, and war." *Journal of Theoretical Politics* 34(1):59–77.