Appendices for "Elections, War, and Gender: Self-Selection and the Pursuit of Victory"

Appendix A: Protocol, Recruitment, Demographics (pp 1 - 5)

Appendix B: Deception Justification (pp 5-7)

Appendix C: Differences in Leader Behavior by Gender (pp 7-9)

Appendix D: Robustness of Effect of NMVW (pp 9-11)

Appendix E: Campaign Messages Analysis (pp 11-15)

Appendix F: Alternative Explanations (pp 16-30)

Appendix References (pp 30-31)

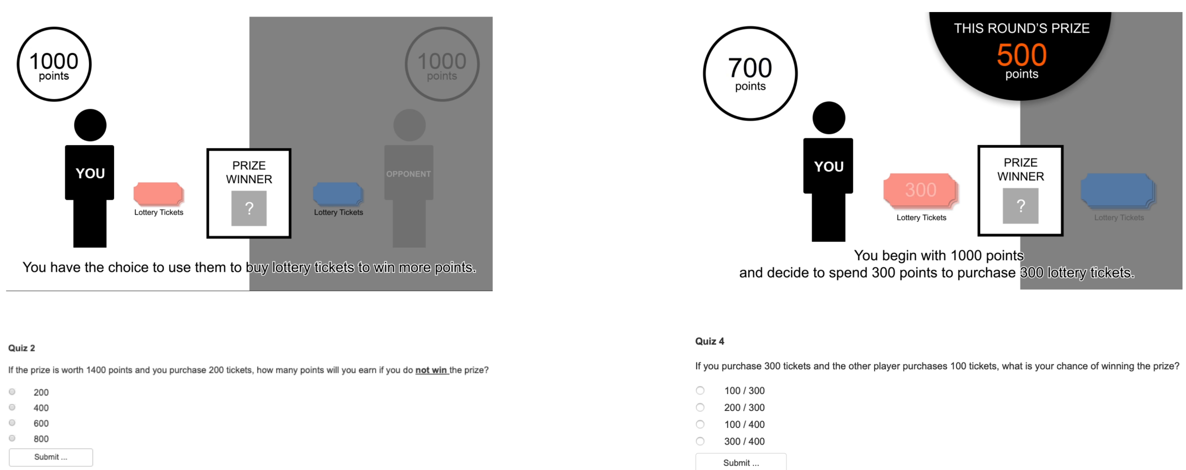
A Appendix: Full Protocol Description, Recruitment, Demographics

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

A.1 Protocol

Before playing, participants watched an animated video explaining the rules of the contest game. We hired a graphics designer to ensure that the instructional video was clear and engaging.[50] Participants then answered four quiz questions to ensure that they understood the mapping between their choices and payoffs. Figure A.1 shows screen shots from the instructional video (top two panels) and from the instructions quiz questions (bottom two panels). If players answered incorrectly, they were shown the correct answer and given an explanation. If players answered correctly, they were told that their answer was correct and given the same explanation.

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



We also included a part of the game after the ICG but before the DSG where participants were randomly sorted into groups and a group leader was randomly chosen. We called this the “Random Selection Game” (RSG). This part was also 12 rounds long with a similar prize value sequence, shown below.[51]

The top two panels of Figure A.2 show screen shots from the game. The top left panel shows an example of a player who was not chosen as group leader and was then asked how many tickets she would have purchased. The screen for a group leader looks similar. The top right panel shows an example of what a player sees at the end of a round. They learned how many tickets that they/their leader purchased, how many tickets the opposing leader purchased, whether they won, and their earnings for that round.[52]

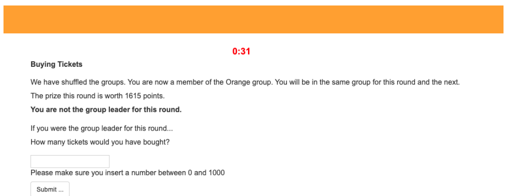
[50]The video is available here: `https://www.youtube.com/watch?v=3ywZvA0CLy8`.

[51]See later section of appendix for analysis replicating ICG results using RSG data.

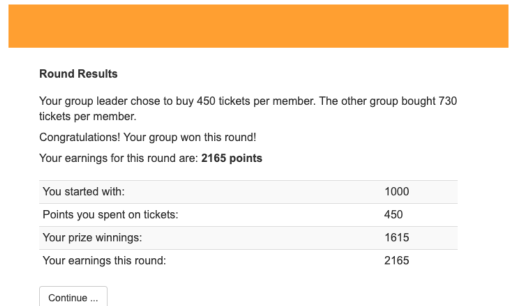
[52]The red times indicate a countdown timer for each decision. This let us drop players who timed-out or dropped out and still keep the game moving. In practice, this rarely happened; respondents did not seem pressed for time.

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

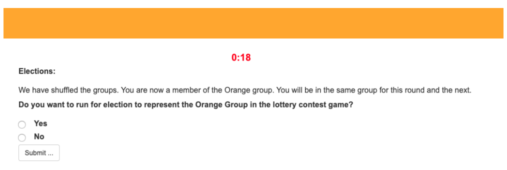
Figure A.2: Screen Captures from the Random Selection and Democratic Selection Parts of the Game



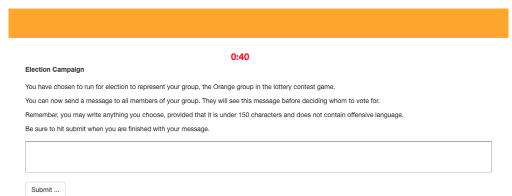
(a) Ticket Purchase



(b) Round Results



(c) Candidate Decision



(d) Campaign Message

Table A.1 shows the prize values for each round and each part of the game. We chose these values to make sure that the prize value was sufficiently enticing to get people to bid, based on test runs. We picked values that weren't exactly round numbers, like 2000, in order to decrease the power of focal points, like bidding exactly one half of the prize value. We varied the prizes and slightly varied the order across sections to minimize order effects.[53]

[53]We created a list of paired prize values that differed from each other by 10 points (eg 2715 and 2705, 1235 and 1225). We then made sure that one value from each pair appeared in the first and second halves of each section (eg, in the ICG, 2715 appears in round 1 - first half - and 2705 appears in round 9 - second half). This makes it very unlikely that participants would identify a pattern and base their play on anticipated future prize values.

Table A.1: Prize Value by Round and Section

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

A.2 Compensation 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.

A.3 Recruitment

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

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.[54] 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.

A.4 Demographics of MTurk Sample

Our sample was closer to the U.S. national averages on most demographic characteristics compared to most university student samples. Figure A.3 shows demographics in our sample compared to the 2018 Cooperative Congressional Election Study (CCES) sample and an in-person university laboratory sample used by Anderson et al. (2013). The CCES uses matching and post-stratification weighting to construct a nationally-representative sample of American adults. We chose Anderson et al. (2013) as a benchmark for student samples because they reported a larger number of demographic characteristics than most studies using student samples. Compared to the CCES, our MTurk participant pool was about 10.5 years younger while the university laboratory pool was about 27.2 years younger. The MTurk sample was much closer to the CCES benchmark in the distribution of education levels. Similarly, our MTurk sample has a more representative distribution of income levels.[55] 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.

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

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

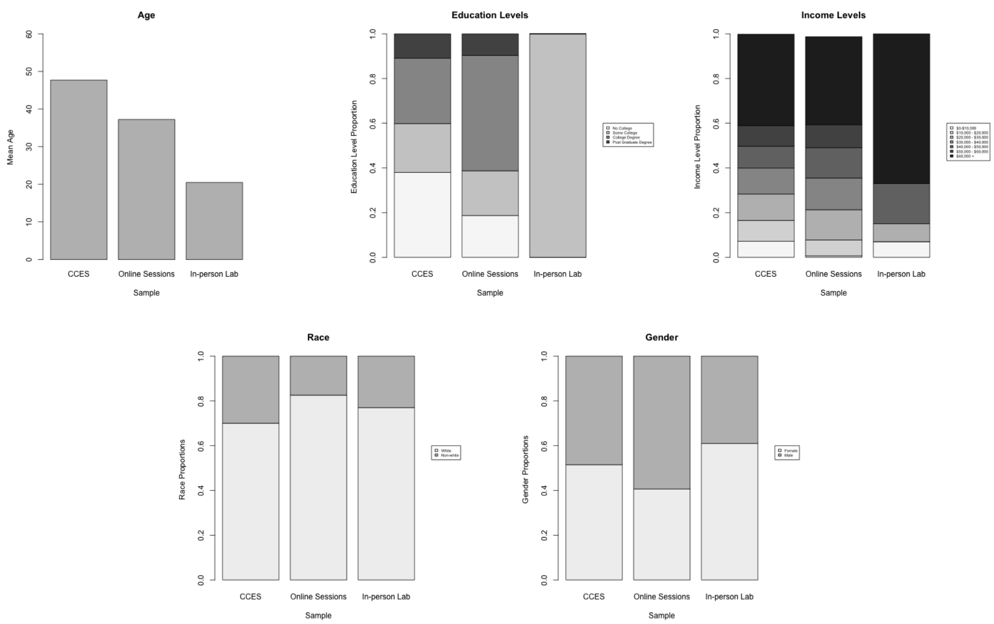


Figure A.3: Comparison of Sample Characteristics

B Appendix: Deception Justification

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

Here, we first describe why neither instance of deception caused harm to the respondents. Then we describe why an alternate protocol would not have been feasible.

B.1 Pre-brief and Debrief

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

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

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

> ***The nature of the study we are conducting required minor deception on our part. Because of the online nature of the experiment, we had to think of ways to handle***

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

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

Lastly, in three of the ten rounds in Part 2, we told all participants that they have been selected as the leader. While leader selection was random in all other rounds of Part 2, it was actually not in those three rounds. We did this in order to make sure we had data on all participants playing the game as the leaders of their groups. This feature did not have any impact on your earnings because we did not select one of these rounds as the one that bonuses are calculated from.

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

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

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

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

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

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

B.3 Full Randomization

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

Figure B.1 shows the binomial distribution when the probability of success (being chosen leader) is 1/8, with 12 draws. For a particular respondent, there is a 20% chance (approximately) that they will never be leader, 35% chance of being leader once, 27% chance of being leader twice, and 18% chance of being chosen 3 or more times.

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

Figure B.1: Binomial Distribution for Random Leader Selection

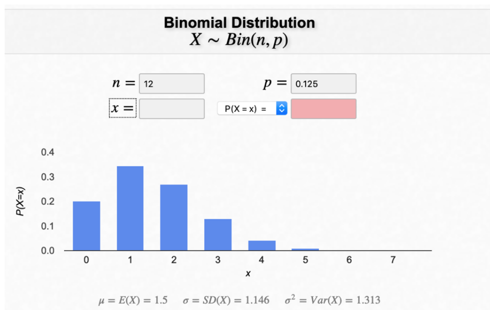


Figure Source: Bognar, Matt. Department of Statistics and Actuarial Science, University of Iowa. `https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html`.

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

C Appendix: Differences in Leader Behavior by Gender

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

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

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

Table C.1: Effect of Gender on Tickets Bought

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

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

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

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

D Appendix: Robustness of Effect of NMVW by Gender

D.1 Leadership and Candidacy

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

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

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

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

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

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

D.2 Winning Candidate

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

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

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

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

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

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

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

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

Table E.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.[56]

The results also obtain if we use a coding of message types based on whether the message contained any amount of a particular type, as opposed to being the message's primary type. Table E.3 replicates Table E.2 using this coding. We again find that, for women, higher NMVWs are

[56]See Appendix E 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 E.2: Effect of Gender and NMVW on Winning, Controlling for Campaign Messages

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

associated with an increased chance of winning, compared to men, even when controlling for this alternate message type coding.

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

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

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

F Appendix: Alternative Explanations

In this section of the appendix, we provide additional discussion and analyses concerning the major alternative explanations laid out in the main manuscript.

F.1 Groups

Here, 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.

All the main arguments in the original manuscript hold when we use the RSG-based measures of NMVW. However, as noted in the main manuscript, data from the RSG tend to emphasize self-selection into candidacy as the primary channel or selection effects.

Figure F.1 shows the distribution of ticket purchases by gender in the RSG. Ticket purchases were very similar for men and women. The variance in ticket purchases was again slightly higher for men than women.

Table F.1 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. Figure F.2replicates Figure 2, illustrating these same claims in terms of estimated effects.

Table F.2 replicates Table 3, showing that greater NMVW in the RSG was associated with writing longer campaign messages, and that this effect was stronger for women than for men. Table F.3 replicates Table E.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.

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

	Dependent variable:					
	Is Leader		Is Candidate		Is Winner	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem.	-1.406^{***}	-2.547^{***}	-1.315^{***}	-2.947^{***}	-0.626^{**}	-0.773
	(0.243)	(0.526)	(0.151)	(0.344)	(0.291)	(0.619)
RSG Nash NMVW	-0.005		-0.031^{***}		0.006	
	(0.013)		(0.010)		(0.014)	
Fem. X RSG Nash NMVW	0.081^{***}		0.117^{***}		0.017	
	(0.022)		(0.016)		(0.026)	
Ave. RSG Tickets		0.003		-0.101^{***}		0.040
		(0.050)		(0.038)		(0.057)
Fem. X Ave RSG Tick.		0.318^{***}		0.458^{***}		0.050
		(0.087)		(0.061)		(0.103)
Constant	-1.574^{***}	-1.615^{***}	-0.095	0.240	-0.728^{***}	-0.893^{***}
	(0.110)	(0.269)	(0.084)	(0.203)	(0.126)	(0.301)
Observations	1,786	1,786	1,786	1,786	645	645
Note:						*p<0.1; **p<0.05; ***p<0.01

Table F.2: Length of Campaign Messages, by Effort and Gender, using RSG data

	Dependent variable:	
	Words in Campaign Message	
	(1)	(2)
Female	-43.855^{***}	-16.635^{***}
	(10.143)	(4.679)
Ave. RSG Tickets	-3.874^{***}	
	(1.024)	
Female 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^{***}	72.290^{***}
	(5.433)	(2.263)
Observations	676	676
R^2	0.036	0.039
Adjusted R^2	0.032	0.035
Residual Std. Error (df = 672)	38.513	38.454
F Statistic (df = 3; 672)	8.362^{***}	9.068^{***}

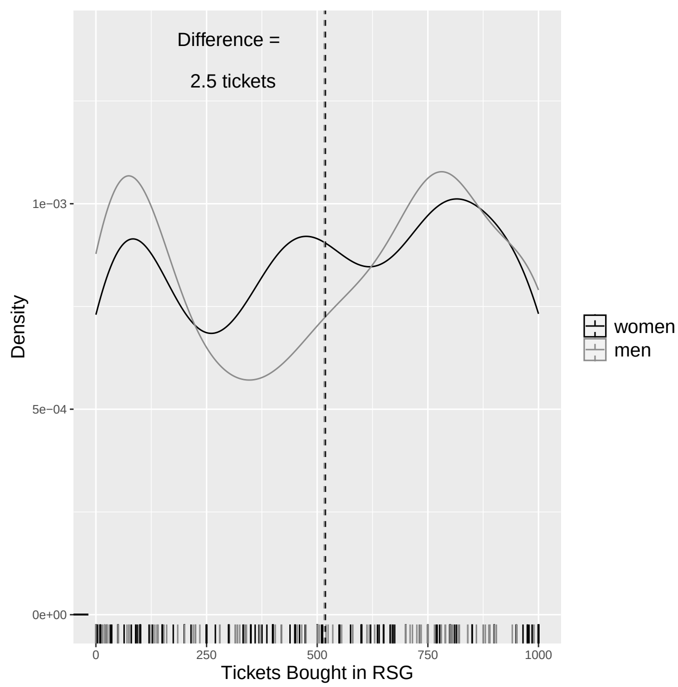
Note: *p<0.1; **p<0.05; ***p<0.01
Averages given in 100s of tickets.

Table F.3: Effect of Gender and RSG NMVW on Winning, Controlling for Campaign Messages

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

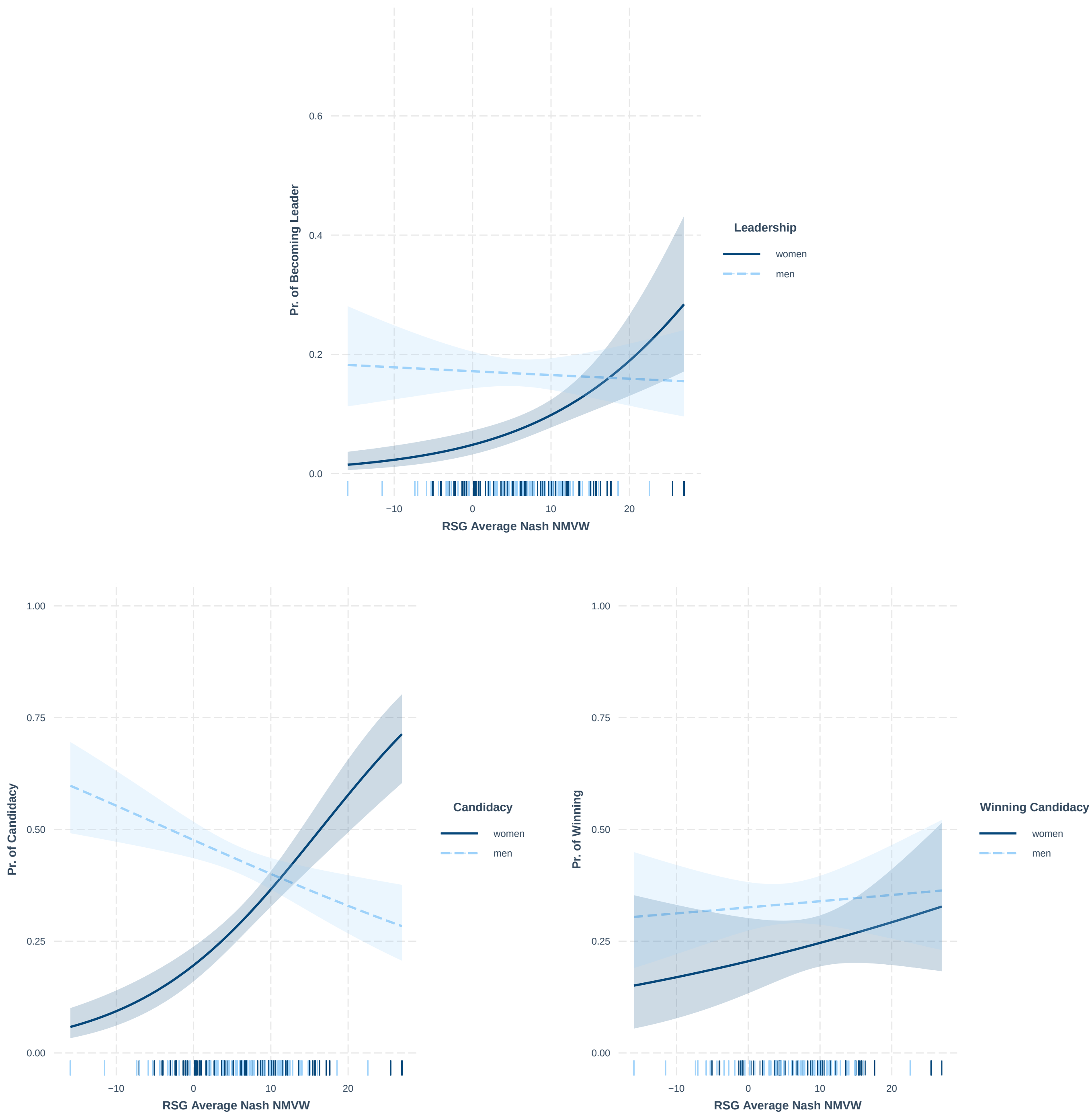
Note: *p<0.1; **p<0.05; ***p<0.01

Figure F.1: Distribution of Tickets Bought in the RSG, 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

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



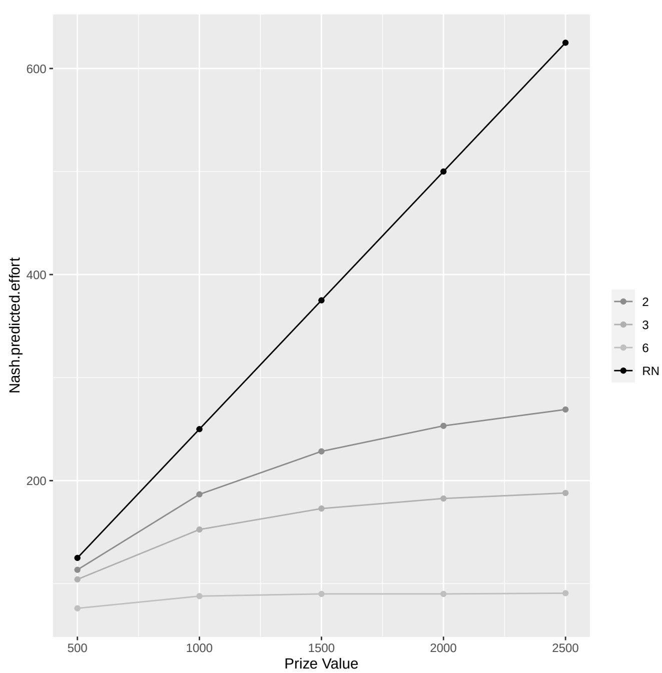
Note: The plot shows predicted probability of being a leader or candidate in a particular round of the DSG. Predictions use estimates from Table F.1.

F.2 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.[57] Empirically, the first effect tends to dominate, so we focus on that effect here.

Baik, Chowdhury and Ramalingam (2020) and [authors] examine how ticket purchases change as we vary the prize value and an individual's degree of risk aversion. Baik, Chowdhury and Ramalingam (2020) analyze the effect of changing budgets for risk averse players. In their Appendix, they derive the first order conditions for ticket purchases for a risk-averse player, using a constant relative risk aversion (CRRA) utility function (Appendix Equation 5). In related work,[58] we took that FOC and used it to numerically calculate optimal ticket purchases as we held a budget constraint fixed, varied the prize value, and varied the coefficient of risk aversion in the CRRA utility function. We reproduced that figure in Figure F.3. The solid, straight line plots the Nash predicted ticket purchase for a risk neutral player. The curved lines plot the same thing, but for increasing coefficients of risk aversion (2, 3, and 6). Higher numbers mean a greater degree of risk aversion. The plot shows the original claim: as we increase the prize value, the marginal effect on ticket purchases is smaller for more risk averse players.

Figure F.3: Nash Effort by Prize Value, as Risk Aversion Varies



For each participant, we calculated their average purchase in the ICG for the low value rounds and the 1225/1235 rounds. We then calculated the change in ticket purchases going from the low

[57] Sahm (2017).
[58] (Chaudoin, Hummel and Park, 2024).

value rounds to the 1225/1235 value rounds.

Table F.4 replicates Table 2 including this measure of risk aversion, as described in the main manuscript. We regressed each of the three outcome variables from Table 2 on the risk aversion measure, our NMVW measure, and their interactions with a female indicator.

We still find that higher-NMVW women were more likely to run, win, and become leaders (bottom interaction terms), even when controlling for risk aversion measures. The risk aversion measures have inconsistent effects for women. They are associated with an increased likelihood of running, but a slightly smaller likelihood of becoming leader.

Table F.4: Effect of NMVW and risk aversion measure on Leadership and Candidacy, by Gender

	Dependent variable:					
	Is Leader		Is Candidate		Is Winner	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	−1.502***	−3.201***	−1.537***	−2.361***	−0.412	−1.927***
	(0.423)	(0.617)	(0.253)	(0.340)	(0.478)	(0.685)
RA Measure	0.0004	0.0004	−0.0001	−0.0001	0.001	0.001
	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.001)	(0.001)
Nash NMVW	−0.003		0.004		−0.007	
	(0.014)		(0.010)		(0.015)	
Ave. ICG Tickets		−0.013		0.015		−0.027
		(0.056)		(0.038)		(0.060)
Female X RA Measure	0.0003	0.0003	0.002***	0.002***	−0.002	−0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female X NMVW	0.099***		0.048**		0.088***	
	(0.030)		(0.019)		(0.034)	
Female X Ave ICG Tick.		0.395***		0.192**		0.352***
		(0.119)		(0.077)		(0.134)
Constant	−1.929***	−1.871***	−0.249*	−0.313*	−0.941***	−0.826***
	(0.216)	(0.266)	(0.148)	(0.185)	(0.261)	(0.303)
Observations	1,748	1,748	1,786	1,786	645	645
Note:						*p<0.1; **p<0.05; ***p<0.01

F.3 Confidence and Learning

F.3.1 Confidence as measured by ICG payoffs

The main manuscript described how we replicated the main table using ICG payoffs as a measure of confidence. Table F.5, which replicates Table 1, shows average ICG payoffs for leaders versus non-leaders, and candidates versus non-candidates.

Table F.5: Differences in ICG payoff, by Gender

	Men		Women	
(I) Leaders vs. Not Leaders				
	leader	*not leader*	*leader*	*not leader*
Ave. ICG Payoff	1401	1308	1377	1310
difference	+93		+67	
(II) Candidates vs. Not Candidates				
	candidate	*not candidate*	*candidate*	*not candidate*
Ave. ICG Payoff	1346	1306	1328	1311
difference	+40		+17	
(III) Winning vs. Losing Candidates				
	winner	*loser*	*winner*	*loser*
Ave. ICG Payoff	1388	1313	1369	1310
difference	+75		+59	

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

Table F.6 regresses 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. However, controlling for this measure of confidence, our NMVW measures still have similar, significantly different effects for men versus women, as in the main manuscript. Even controlling for our proxy for confidence, the significance of the average ICG tickets purchased, our measure for NMVW, persists.

We then replicated that analysis again focusing on the likelihood of winning an election, conditional on selecting into candidacy, in Table F.7. 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 the same effects on the probability of winning for men and women candidates.

Table F.6: Effect of ICG Payoff on Candidacy, by Gender

	Dependent variable:		
	Candidate		
	(1)	(2)	(3)
Female	−0.013	−2.690***	−1.126**
	(0.489)	(0.656)	(0.527)
Ave. ICG Payoff	0.001**	0.001**	0.001**
	(0.0002)	(0.0002)	(0.0002)
Ave. ICG Tickets		0.022	
		(0.031)	
Nash NMVW			0.006
			(0.008)
Female × Ave. ICG Payoff	−0.0004	0.0001	0.0001
	(0.0004)	(0.0004)	(0.0004)
Female × Ave. ICG Tickets		0.364***	
		(0.058)	
Female × Nash NMVW			0.091***
			(0.014)
Constant	−1.039***	−1.204***	−1.108***
	(0.320)	(0.395)	(0.334)
Observations	1,786	1,786	1,786
Log Likelihood	−1,171.483	−1,135.313	−1,135.313
Note: Averages given in 100s of tickets.	*p<0.1; **p<0.05; ***p<0.01		

Additionally, for confidence in abilities to explain our results, boosts to confidence - stemming from a higher payoff in one round of the ICG - should be associated with a higher ticket purchase in ensuing rounds *and* this effect should differ by gender. We see neither relationship. Table F.8 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.

F.3.2 Confidence as measured by electoral wins

Confidence might also refer to one's beliefs about their ability to win elections. This facet of confidence is inherently relational, since belief in one's ability to win elections implies a belief that one can win an election relative to other possible candidates. Pruysers and Blais (2017) and Fox and

Table F.7: Effect of ICG Payoffs on Electoral Success, by Gender

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

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

Table F.8: Effect of Lagged Payoffs on Ticket Purchases, by Gender

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

Lawless (2011) link this aspect of confidence with gender differences in election aversion. However, Bernhard and de Benedictis-Kessner (2021) do not find that women candidates are especially discouraged by electoral losses.

We checked to see if boosts to electoral confidence (from winning election), affected men and women differently. We do not find evidence of this.

Table F.9, 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 $t - 1$.[59] 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 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.

F.3.3 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

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

Table F.9: Effect of Electoral Wins/Losses on Candidacy/Winning, by Gender

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

whether this also differs between those who are eventually an elected leader and those who are not elected leaders.

Figure F.4 shows data from the ICG. In each round, for each participant, we calculated the number of tickets they purchases 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 F.4 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.

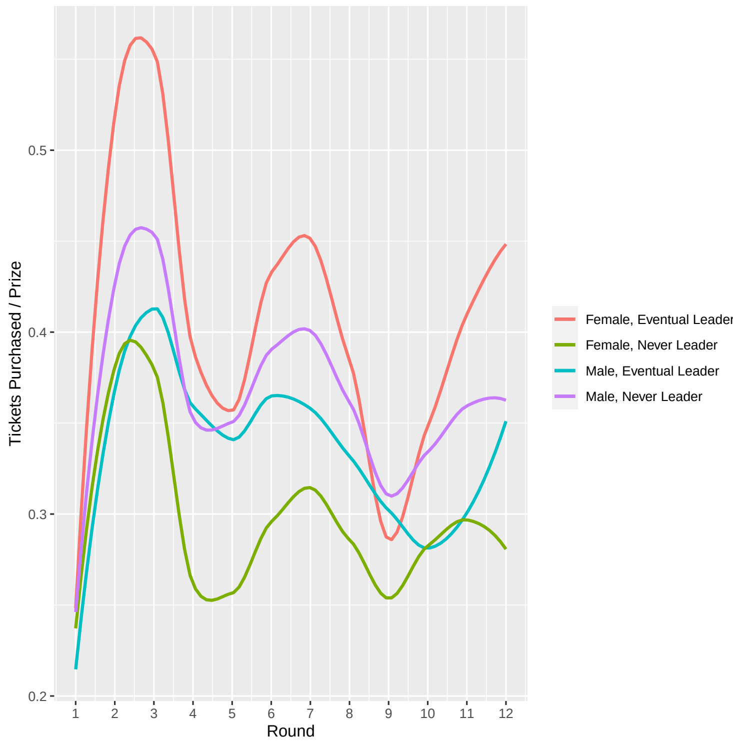
First, the patterns match those highlighted in the main text. Eventually-elected women tend to have the highest ticket purchases across rounds, followed by both elected and never-elected men, and then women who were never elected. Second, there does not appear to be differential changes in behavior over 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 F.10 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 learning from observed play between groups. Columns 3 and 4 repeat the same process for those who eventually do and do not become candidates.

F.4 Election Effects

The DSG portion of the game adds an election, which potentially induces two other changes that are distinct from selection effects. The experience of being elected can affect behavior directly (Chaudoin, Hummel and Park, 2024). This "election effect" could vary across gender. We consider

Figure F.4: Learning over time in the ICG, by gender and eventual leadership



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.

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 with the code in the replication file; omitted for length).

F.5 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),

Table F.10: Learning over time, by gender and eventual leadership/candidacy

	Eventual Leaders	*Never Leaders*	*Eventual Cand.*	*Nev. Cand.*
	boughtTickets			
	(1)	(2)	(3)	(4)
Female	53.274	-101.819^{**}	1.291	-120.420
	(49.347)	(48.890)	(39.902)	(74.974)
Round number	2.605	3.759	2.957	4.096
	(3.992)	(4.529)	(3.256)	(7.938)
Female X Round number	1.784	-1.954	1.946	-6.054
	(6.705)	(6.643)	(5.422)	(10.187)
Constant	532.404^{***}	553.450^{***}	537.884^{***}	566.249^{***}
	(29.378)	(33.331)	(23.963)	(58.419)
Observations	948	852	1,464	336
Note:			*p<0.1; **p<0.05; ***p<0.01	

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. (Code for this is available in the replication materials. Results omitted for length.)

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.

Baik, Kyung Hwan, Subhasish Chowdhury and Abhijit Ramalingam. 2020. "The Effects of Conflict Budget on the Intensity of Conflict: An Experimental Investigation." *Experimental Economics* 23:240–258.

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, Sarah Hummel and Yon Soo Park. 2024. "The Election Effect: Democratic Leaders in Inter-Group Conflict." *International Studies Quarterly* 68(1):sqad107.

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* .

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

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

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