Appendices

Leaders but not Authorities: Gender, Veterans, and Messages About National Security

Appendix A Principles and Guidance for Human Subjects Research

The studies conducted were approved by Institutional Review Boards, and were deemed to be of minimal risk. In what follows, we describe our research approach within the context

of American Political Science Association's Principles and Guidance for Human Subjects Research, as outlined here. We address each relevant component of the guidelines:

- **Power:** APSA guidelines note the awareness of potential power differentials. Our studies are not conducted vulnerable participants and in vulnerable communities. Nonetheless, we took steps in our study design to ensure that participants never felt coerced into completing the study. First, we worked with a survey company for our main studies, which means we never directly interacted with study participants. In conducting our post-hoc studies during the review process, we posted our studies via CloudResearch – which participants could choose to participate if they wanted to. Next, participants in all studies could skip any items they wished in the study, and would still obtain the reward (via either the survey company or through the convenience sample platform). Finally, all our studies were fully de-identified, as participants were never asked for any identifying information.

- **Consent:** As APSA guidelines note, the studies in this manuscript included informed consent. All participants received an informed consent screen (as these studies were online) and only those who agreed to participate continued on to the rest of the study. Moreover, as we were working with survey companies, we note that there were two layers of consent. In our main studies, first, participants consented to be part of a survey panel (i.e. allow the survey company to retain their information); then they consented to being part of our specific study. In our post-hoc checks participants first decided to join the convenience sample platform, and then opted to consent to be in our studies.

- **Deception:** Our study depends on fictitious politicians; participants in our study were debriefed – in accordance with Institutional Review Board guidelines – about the fictitious politicians, as well as the goals of the study. In the post-hoc checks, participants were told in advance that these are hypothetical politicians. In addition, we also explained to the participants why we relied on fictitious politicians in the first place. In accordance with the IRB review, it was determined that this study had minimal risk of harm, which justified using fictitious candidates. Moreover, all the actual facts in the treatments of the main studies were truthful. We note that aside from the fictitious candidates, we were clear about our identities as researchers, our institutions and the future use of the data.

- **Harm:** In our assessment, the potential for harm as a result of participation in out studies is low. We use unobtrusive survey measures, and there are no social or economic harms that could result from these studies.

- **Confidentiality:** We take a number of steps to ensure confidentiality. The survey software we use has confidentiality protections, and we do not ask any questions that may lead to identifying information. As APSA suggests, we inform participants of the steps we've taken to ensure confidentiality in the consent statement.

- **Impact:** APSA guidelines state "In general, political science researchers should not compromise the integrity of political processes for research purposes without the con-

sent of individuals that are directly engaged by the research process." Indeed, the design of our studies does not involve real election scenarios or real candidates seeking public office – as manipulating information about these candidates could affect political outcomes outside our research context. Therefore, our impact is the goal of greater understanding about the role speaker gender plays in individual response to speech.

- **Compensation for Participants:** Both our main studies are fielded through the survey company Dynata (though, at the time of fielding the company was called SSI). Participants are compensated with points determined by the survey company based on the length of study. Participants can then turn these points into either gift cards or other financial rewards. Participants agree to this compensation approach when the join the Dynata panel. As it is the survey company's goal to retain participants in the panel, they ensure that the points awarded are appropriate. Our post-hoc checks are run via Mechanical Turk via CloudResearch. In this case, participants were paid to ensure they were making above minimum wage based on the timing of the studies.

Appendix B Overview of Additional Studies

As this manuscript relies on several pre-tests and post-hoc checks, we present the full overview of this additional data in Table B.1. In this table, we also note which studies were done omnibus. Finally, we also note where in the SI the results of these studies appear.

Appendix C Image Tests

To ensure there were no confounding factors in the images used – for example, participants believing that one politician was older or more attractive than the other – we used images that were tested prior to the fielding of this study for age and attractiveness. We also conducted additional testing of the images during the review process.

C.1 Characteristics

In the pre-test, participants (N=171, MTurk), were randomly assigned to either the man or the woman from our studies. They were shown *only* the image, and there was no accompanying text – no name, no additional statements.[35]. Participants rated the candidate's attractiveness on a scale of 0 (least attractive) to 5 (most attractive) and friendliness (0 (least friendly) to 5 (most friendly).

[35]More participants were recruited in the original study, but only this group was randomly assigned to rate photographs at the end of the study; the other participants engaged in unrelated tasks. Among our participants, in addition, there were 6 participants who were assigned to this treatment whose IP were not located in the US; these participants are excluded which leaves us with N=171. The results do not change if these 6 participants are included: attractiveness difference is 0.13 ($p = 0.45$); friendliness is 0.2 ($p = 0.20$)

Table B.1: Pre-Tests and Post-Hoc Checks

	Content	N	SI	Omnibus?
Pre-Test				
Image Rating	Participants evaluate the attractiveness and friendliness of the politician; includes *only* the image	N=169	SI C	Stand-alone
Veteran Check	Tests difference in responses to "combat veteran" vs. "veteran"	N=149	Ftnote 15	Stand-alone
Post-Hoc Checks				
Flag Test 1	Participants rate politicians in images with different-sized flags; ratings are of the candidates	N=397	SI C	With Typicality Check 1, Veteran Combat Perception
Flag Test 2	Participants rate politicians in images with different-sized flags; ratings are of participants' nationalism levels	N=347	SI C	With Article Topic Study 1
Veteran Combat Perception	Participants rate whether a man or a woman veteran were more likely to experience combat	N=397	SI E	With Typicality Check 1, Flag Test 1
Article Topic, Study 1	Participants read the study from Study 1, and answered what the topic was	N=347	SI D	With Flag Test 2
Typicality Check 1	Participants saw politician images and evaluated how likely this person would be to increase defense spending	N=397	SI E	With Flag Test 1, Veteran Combat Perception
Typicality Check 2	Participants saw politician images and evaluated how likely this person would be to increase defense spending	N=351	SI E	Stand-alone
Statement Expectations	Participants saw statements from Study 1 and rated the topic, expectancy	N=599	SI E	Stand-Alone

Table C.1: Comparisons of Images

	Man Mean	Woman Mean	p-value (two-tailed)
Attractiveness (N=164)	2.73	2.65	0.65
Friendliness (N=169)	3.16	3.33	0.28

Results are from a t-test comparing friendliness and attractiveness by treatment.

C.2 Post-Hoc Image Tests

We also conducted two additional tests of our images during the review process (Test 1: MTurk via CloudResearch N=397; Test 2: MTurk via CloudResearch N=347). Both tests focused on the size of the flag behind the politician in our images. In these checks we varied the size of the flag behind the politician; test 1 looked at perceptions of the politician's leadership, patriotism, competence and favorability. Test 2 looked at the extent to which the size of the flag in the images affected the *participants' own* levels of nationalism. In each test, our participants were assigned to one of four conditions: (1) Man with a large flag, (2) Woman with a large flag, (3) Man with a small flag, (4) Woman with a small flag.

C.2.1 Image Test 1: Evaluations

In the first test we tested participants' evaluations of the politician's leadership, patriotism, competence and favorability after exposure to the image using competence and patriotism questions that are **not** measures that we include in our main studies. We present our results in Table C.2 as coefficients, with "Man/Large Flag" as the excluded group. The size of the flag does not seem to have any significant effects on ratings relative to the excluded group. Of particular note is the comparison to the woman with the smaller flag: there are no significant differences, and also no clear substantive patterns in the direction of the coefficient. We also conduct the analyses with cases that have duplicate latitude/longitude locations and find very similar results.

We can also make another comparison: would the woman have gotten different ratings with the larger flag rather than the smaller flag? We see no evidence to this point on any one of our outcome measures (results also robust to exclusion of duplicate location cases). The differences in ratings between a woman with the large flag and one with a small flag are as follows (t-test, two-tailed p-values): Leadership = 0.18 (p=0.25); Patriotism = -0.12 (p=0.60); Competence = -0.04 (p=0.73); Favorability = 0.002 (p=0.98).

For candidates who are men (t-test, two-tailed p-values): Leadership = 0.17 (p=0.25); Patriotism = 0.11 (p=0.66); Competence = -0.10 (p=0.34); Favorability = 0.04 (p=0.77).

Table C.2: Test 1: Effect of Flag on Politician Evaluations

	Coefficient (Std. Err.)	**Coefficient (Std. Err.)**	**Coefficient (Std. Err.)**	**Coefficient (Std. Err.)**
	Leadership (1-7 scale)	Patriotism (1-10 scale)	Competence (1-5 scale)	Favorability (1-7 scale)
Woman/large flag	-0.070	-0.013	-0.122	0.213
	(0.153)	(0.241)	(0.109)	(0.144)
Man/small flag	0.169	0.106	-0.100	0.041
	(0.152)	(0.239)	(0.108)	(0.143)
Woman/small flag	0.111	-0.134	-0.160	0.211
	(0.152)	(0.239)	(0.108)	(0.143)
Constant	5.101**	7.384**	2.040**	4.929**
	(0.108)	(0.169)	(0.077)	(0.102)
N	396	396	397	397

* p<0.05, ** p<0.01. All models are OLS. Excluded category is "large flag, man"

C.2.2 Test 2: Participants' Nationalism

In Test 2 we consider if our treatment (the image) affects the participants' *own* level of nationalism using measures from Huddy and Khatib (2007). The nationalism measure asks participants how often they use "we" rather than "they" when talking about Americans. The response options range from "very frequently" to "very rarely." The patriotism measure asks participants how good it makes them feel to see the American flag flying; response options range from "very good" to "very bad."

Results are in Table C.3; the man with the large flag is the excluded category. We do not see any evidence that the size of the flag affects the respondents' nationalism or patriotism.

Table C.3: Test 1: Effect of Flag on Feelings of Nationalism or Patriotism

	Coefficient (Std. Err.)	**Coefficient (Std. Err.)**
	Nationalism (1-4 scale)	Patriotism (1-5 scale)
Woman/large flag	0.050	-0.183
	(0.145)	(0.146)
Man/small flag	-0.079	0.139
	(0.145)	(0.145)
Woman/small flag	-0.125	0.096
	(0.145)	(0.146)
Constant	2.136**	2.125**
	(0.102)	(0.103)
N	347	347

* p< 0.05, ** p< 0.01 OLS models. Nationalism is on 1 to 4 scale, patriotism is on a 1 to 5 scale

Appendix D Study 1

D.1 Study 1: Additional Checks

D.1.1 Study 1: Perceptions of Topic

During the review process, we considered whether the treatment used in Study 1 may have primed the importance of an issue *besides* defense. The treatment was in the headline, but the final lines of our treatment included some economic terms; a reviewer posed the possibility that people who read the treatment more carefully were *less* likely to think it was about defense. Therefore, we tested whether participants assigned to the defense treatment, did perceive the treatment as one about defense, and whether perceptions about our treatment differed by attention levels. We fielded another post-hoc study to check on the treatment used in Study 1 (N=347, MTurk recruited via CloudResearch).

In this post-hoc check, we randomly assigned our participants to either the defense or neutral treatment from Study 1. We also measured the amount of time participants remained on the screen. After exposure, we then asked participants what the story they read was about; our outcome measure was list of topics (presented randomly) which included both defense and economic measures.

We find the following set of patterns:

- Of the participants assigned to the defense treatment, 87.3% report that the article is about defense, compared to 8.05% in the neutral version.
- Among participants assigned to the defense article who are above median on the time spent reading the article, a higher percentage report that it is about defense: 90.70%, relative to 83.9% among those at median time and below. We see similar results when we estimate a model using categorization as a dependent variable and time as an independent variable: participants who spent more time were more likely to list defense.

D.1.2 Study 1: Checks on Balance

In Study 1, due to a Qualtrics error more participants were assigned to the conditions with women (approx. 75% of the sample) than the conditions with men. We conducted checks to ensure that despite the issue, the groups were still similar. We present two different checks in Table D.1.

D.1.3 Study 1: Additional Results

In the main text, Figure 2 presents the results of a model that estimates the interaction between our experimental factors. We present coefficient estimates below in Table D.2, as both probit and OLS estimates (Figure 2 in the text presents the estimates from the OLS).

D.1.4 Study 1: Cross-Partisan and Partisan Match

Since we do not have any a priori expectations about partisanship, we can also consider our results treating partisanship as a nuisance factor maintaining a higher N per condition. We

Table D.1: Randomization Checks

	All Treat			Cand. Gender		
Variable	**Coefficient**	**(Std. Err.)**	**p-value**	**Coefficient**	**(Std. Err.)**	**p-value**
Ind. Participant	0.164	(0.150)	0.276	-0.029	(0.031)	0.358
Repub. Participant	0.102	(0.168)	0.542	-0.031	(0.035)	0.377
Female Participant	-0.190	(0.131)	0.147	0.025	(0.027)	0.364
Constant	3.32	(0.130)	p<0.0001	0.757	(0.027)	p<0.0001

OLS regression using either (1) the full treatment assignment as dependent variable or (2) assignment to either woman/man condition; N=1,025

Table D.2: Coefficient Estimates, Study 1, Effect of Treatment on Defense Importance

	Probit Coef SE	OLS Coef SE
Woman	-0.202	-0.030
	(0.287)	(0.051)
Democrat	-0.031	-0.005
	(0.295)	(0.054)
Woman× Democrat	0.411	0.068
	(0.368)	(0.066)
Defense	0.626**	0.151**
	(0.298)	(0.062)
Woman×Defense	-0.377	-0.113
	(0.369)	(0.073)
Democrat × Defense	-0.185	-0.057
	(0.401)	(0.082)
Woman × Democrat × Defense	-0.075	0.018
	(0.482)	(0.096)
Constant	-1.132***	0.094**
	(0.218)	(0.040)
N	1029	1029

** < 0.05; *** < 0.01; OLS estimates produce Figure 2. Outcome is a binary defense importance variable.

also consider patterns if participants are assigned to either their own party or the opposing party. We present the results below as well.

Table D.3: Treating Partisanship as a Nuisance Factor

Variable	Coefficient	(Std. Err.)
Woman	0.013	(0.033)
Defense	0.116***	(0.041)
Woman × Defense	-0.107**	(0.047)
Constant	0.091***	(0.027)
N	1029	

** p < 0.05; *** p< 0.001. OLS model using binary importance of defense measure.

Table D.4: Partisanship as In and Out-Party

Variable	Coefficient	(Std. Err.)
Woman	-0.016	(0.053)
In-Party	-0.029	(0.059)
Woman × In-Party	0.039	(0.071)
Defense	0.10	(0.063)
Woman × Defense	-0.037	(0.073)
In-Party × Defense	0.023	(0.091)
Woman × In-Party × Defense	-0.084	(0.105)
Constant	0.10**	(0.026)
N	840	

** p < 0.05; *** p< 0.001. OLS model; binary importance of defense measure. Excludes non-leaning independents/participants who did not answer the party question.

D.1.5 Study 1, Additional Outcome Measures

In Study 1, we also included additional outcome measures (similar to those in Study 2). We had no a priori hypotheses as to how the patterns on these outcome variables should differ from leadership. As the patterns are generally the same across all outcome variables, and as Study 1 serves as a preliminary study (with notable limitations), we limit the amount of space it takes up in the final manuscript and report the additional outcome measures in Table D.6 as means/t-tests.

Table D.5: Change in Believing Candidate is a "Strong Leader," Study 1, by in vs. out-party

	Mean Evaluation, Neutral	Mean Evaluation, Threat/Defense	Change Due To Treatment
By Gender/Party:			
In-Party Woman	4.25	4.54	+0.29 (p=0.068)
In-Party Man	4.72	4.83	+0.10 (p=0.86)
Politician Gender Diff.	+0.48 (p=0.014)	+0.28 (p=0.27)	
Out-Party Woman	3.75	4.16	+0.40 (p=0.01)
Out-Party Man	3.96	4.26	+0.30 (p=0.23)
Politician Gender Diff.	+0.20 (p=0.36)	+0.10 (p=0.67)	

Each comparison is a two-tailed t-test. Outcome measure: mean belief that a politician is a strong leader, 1 to 7 scale. N=820, but more participants were randomized to the conditions with women; N is lower because pure independents, participants who didn't answer the party question, are excluded

Table D.6: Change Due to Treatment, Additional Outcome Measures

	Mean Evaluation, Neutral	Mean Evaluation, Threat/Defense	Change Due To Treatment
Favorability:			
Woman	4.12	4.25	+0.13 (p=0.20)
Man	4.35	4.27	-0.08 (p=0.64)
Politician Gender Diff.	+0.23 (p=0.08)	+0.04 (p=0.78)	
Vote:			
Woman	3.74	3.95	+0.21 (p=0.09)
Man	4.10	3.94	-0.16 (p=0.44)
Politician Gender Diff.	+0.36 (p=0.02)	-0.01 (p=0.94)	

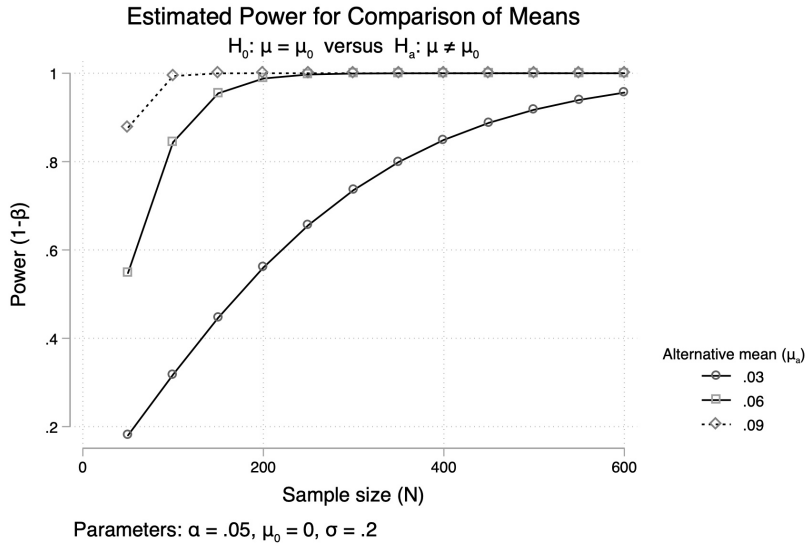
Each comparison is a two-tailed t-test. Outcome measure: mean favorability, on a 1 to 7 scale, mean likelihood of voting on a 1 to 7 scale. Favor, N=1,025; vote, N=1,024, more participants were randomized to the conditions with women

Appendix E Study 2

E.1 Study 2: Power Analysis

To consider power in Study 2, we rely on the effect sizes in Study 1. In conditions where participants see women of their own party (which is what participants receive in Study 2), the 95% confidence interval around the effect size on a binary dependent variable is (-0.090, 0.058); the standard deviation is 0.04. We use these parameters to calculate how much power we would have at various sample sizes. Our results (Figure E.1) suggest that we have large enough groups in Study 2 to observe group differences based on the confidence intervals observed in Study 1 with an $\alpha = 0.05$ and a power higher than 0.80.

Figure E.1: Power, Based on Study 1 Results



E.2 Study 2: Sample Characteristics

We use pre-treatment demographic measures to consider the characteristics of our sample in Table E.1. Our sample was recruited by RNSSI, since rebranded as Dynata.

Table E.1: Sample Characteristics, Study 2

Democrats (includes leaners)	46.56%
Republicans (includes leaners)	38.24%
Mean ideology (full sample, higher values = liberal)	4.09
Women Participants	50.24%
White	76.59%
African American	15.46%
Modal Education Categories	High school (25.3%)
	Some College (25.2%)
Income (modal category)	$50,001 to $75,000 (19.5%)

E.3 Study 2: Attention Check

We conduct a randomization check by predicting treatment assignment by partisanship, which was asked pre-treatment. Other demographics were asked post-treatment following Klar, Leeper and Robison (2020). This produced a χ^2 of 5.49 (p-value of 0.600).

Our study included a manipulation check to see whether the respondent recalled the gender of the candidate presented in the treatment: asking participants, " Earlier you were asked to read a piece of text about a congressional candidate. What was the candidate's gender?" The response options are: (a) Man, (b) Woman and (c) Don't remember. Across the whole sample 87.07% pass the check (e.g. get the gender correct), 12.03% fail (e.g. get the gender wrong) and 0.9% report not remembering the gender.

We then track whether we see any evidence that failure rates were differential by treatment in Table E.2.

We see little evidence of any particular patterns, with predicted passage rates by group shown in Table E.3.

E.4 Study 2: Post-Hoc Checks

During the review process, we conducted three additional checks on our treatments. We present all the tests below.

E.4.1 Post-Hoc Check 1, Study 2: Partisan Expectations

We ran two versions of Post-Hoc Check 1: in both, we randomly assigned participants to one of four conditions: (1) Democratic Man, (2) Democratic Woman, (3) Republican Man,

Table E.2: Predicting Attention Check Passage (1=passed, 0 failed, did not remember)

Variable	**Coefficient**	**(Std. Err.)**
Woman Candidate	-0.353	(0.302)
Vet	-0.027	(0.319)
Woman × Vet	0.010	(0.425)
Statement	-0.409	(0.298)
Woman × Statement	0.215	(0.404)
Vet × Statement	0.316	(0.432)
Woman × Vet × Statement	0.243	(0.592)
Constant	2.151**	(0.225)
N	1,663	
χ^2	7.64 (p=0.365)	

Logit, binary outcome where: 1=passed, 0 failed, did not remember; * p< 0.05 ** p< 0.01

Table E.3: Predicted Passage Rates Using Estimates from Table E.2

Woman, Vet, Statement	0.895
Woman, Vet, No Statement	0.856
Woman, Not Vet, Statement	0.832
Woman, Not Vet, No Statement	0.857
Man, Vet, Statement	0.884
Man, Vet, No Statement	0.893
Man, Not Vet, Statement	0.851
Man, Not Vet, No Statement	0.895

Predicted passage rates based on the model in Table E.2

(4) Republican Woman. We used the same images as those used in the main study and participants were asked to rate how likely this person would be to increase spending on defense. The goal was to track whether people saw men and women as having different views on defense *regardless* of the partisan cue. The difference between our two versions of this check is whether the politicians were described as veterans (N=397, MTurk via CloudResearch) or did not have that additional description (N=351, MTurk via Cloud Research). We report our results in Table E.4.

Table E.4: Post-Hoc Checks: Partisan Expectations on Defense Spending

	No Veteran Coef (SE)	**Veteran Coef (SE)**
Politician = Dem	-2.047 (0.276)**	-1.338 (0.242)**
Politician = Woman	-0.033 (0.276)	-0.034 (0.234)
Pol PID × Pol Woman	-0.323 (0.394)	0.126 (0.338)
Constant	7.286 (0.193)**	7.94(0.168)**
N	351	397

* p< 0.05, ** p< 0.01. OLS. Dependent variable: likelihood of increasing defense spending on a scale of 1 to 10 (higher = more)

E.4.2 Post-Hoc Check 2, Study 2: Statement

We presented a sample of N=599 participants (MTurk via Cloud Research) with the statements the candidates made in Study 2. Participants were randomly assigned to either the statement the Democratic candidate made or the statements the Republican candidate made, and informed of the partisanship of the candidate making the statement just as in Study 2. Then, participants were asked if they perceive the statement as expected or unexpected for the candidate (7-point response scale). Using a Kolmogorov-Smirnov test we find no difference between the groups; combined KS-D of 0.058, p=0.691.

E.4.3 Post-Hoc Check 3, Study 2: Expectations of Men and Women Who Are Veterans

We also consider whether people assume that men and women who are veterans had different military experiences. We randomly assigned participants to one of four conditions: (1) Woman Veteran, own party, (2) Man Veteran, own party, (3) Woman Veteran, other party, (4) Man Veteran, other party. The images used were identical to those in Studies 1 and 2.

After the exposure, participants answered the following set of questions: (1) How much active combat do you think [candidate] has experienced? (2) How many dangerous situations did [candidate] experience while serving in the military? (3) How likely or unlikely do you think it was that [candidate] was wounded while serving in the military?

The first two measures had a set of response options on a 4-point scale, ranging from Very Many/A Great Deal to Almost None. The third measure is on a 5-point scale, ranging from Extremely Likely to Extremely Unlikely. The full results are in Table E.5.

Table E.5: Post-Hoc Checks: Perceptions of Women Veterans

	Own Party Man−Woman Difference	**Other Party Man−Woman Difference**
Combat (1-4 scale)	0.04 ($p = 0.66$)	-0.07 ($p = 0.59$)
Danger (1-4 scale)	0.05 ($p = 0.60$)	0.01 ($p = 0.92$)
Wounded (1-5 scale)	-0.19 ($p = 0.12$)	0.03 ($p = 0.86$)

Results are based on a series of t-tests which compare by gender, within party. Reported p-values are two-tailed.

The results suggest that there are no differences in perceptions of men and women who are veterans *within party.* We do, however, see differences across party – people are more likely to believe that members of their own party have valor.

E.5 Study 2: Coefficient Estimates, Leadership, Favorability, Qualification

We present our results as figures in the main manuscript and coefficient estimates here. All the models used for the estimates in the text do not include controls; we present results with controls as a check below. An additional check following our pre-registration is excluding participants who failed the memory check (see Table E.2). We present the results for the leadership variable (Table E.6), favorability (Table E.7) and qualifications (Table E.8). The patterns for knowledge are shown in the next section of the SI. In these analyses, we treat statement as a nuisance factor and exclude it from the model.

We also present the marginal effect of the candidate being a combat veteran across outcome measures (Table E.9). We find that for both men and women, the effect of being a combat vet is always positive. Focusing on the marginal effect of gender – i.e. whether being described as a woman or man matters for leadership perceptions – we see *very little* evidence of any gender differences.

Table E.6: Effect of Treatment on Leadership Ratings (OLS)

	In-Text		Controls		Indep. Excl.	
	Coeff	**(SE)**	**Coeff**	**(SE)**	**Coeff**	**(SE)**
Woman Candidate	-0.135	0.104	-0.107	0.105	0.062	0.115
Vet	0.161	0.104	0.167	0.105	0.180	0.114
Woman × Vet	0.121	0.147	0.088	0.147	0.036	0.161
Ideology			-0.038	0.022		
Woman Partic.			-0.186*	0.080		
White			-0.221*	0.094		
Education			-0.039	0.029		
Income			0.015	0.018		
Independent			-0.221*	0.104		
Intercept	4.446**	0.073	4.959**	0.181	4.452**	0.081
N	1,643		1,599		1,389	

*p<0.05, ** p<0.01; OLS with leadership dependent variable; "in text" column produces Figure 4 and Figure 5.

Table E.7: Effect of Treatment on Favorability Ratings (OLS)

	No Controls		Controls		Indep. Excl.	
	Coeff	**(SE)**	**Coeff**	**(SE)**	**Coeff**	**(SE)**
Woman Candidate	0.043	0.092	0.081	0.093	0.079	0.105
Vet	0.224*	0.092	0.205*	0.093	0.243*	0.105
Woman × Vet	-0.045	0.131	-0.057	0.131	-0.073	0.148
Ideology			-0.078**	0.020		
Woman Partic.			-0.108	0.071		
White			-0.102	0.083		
Education			-0.045	0.026		
Income			0.045**	0.016		
Independent			-0.208*	0.092		
Intercept	4.163**	0.065	4.581**	0.160	4.183**	0.074
N	1,659		1,612		1,402	

*p<0.05, **p<0.01 OLS estimate; favorability dependent variable.

Table E.8: Effect of Treatment on Qualification Ratings (OLS)

	No Controls		Controls		Indep. Excl.	
	Coeff	**(SE)**	**Coeff**	**(SE)**	**Coeff**	**(SE)**
Woman Cand.	0.047	0.062	0.063	0.063	0.051	0.069
Vet	0.205**	0.062	0.205**	0.064	0.181**	0.069
Woman × Vet	-0.045	0.088	-0.071	0.089	-0.021	0.098
Ideology			-0.027	0.014		
Woman Partic.			-0.021	0.048		
White			-0.091	0.057		
Education			0.007	0.018		
Income			0.016	0.011		
Independent			-0.138*	0.063		
Intercept	3.327**	0.044	3.369**	0.109	3.329**	0.049
N	1,660		1,614		1,401	

*p<0.05, **p<0.01, OLS, qualification as dependent variable.

Table E.9: Marginal Effects on Outcome Measures E.2

	Marginal Effect of Veteran Status:		
	Change for Men	Change for Women	Location
On Leadership	0.16 (p=0.121)	0.28(p=0.007)	In-Text
			Figure 4a
			Coeffs. in Table E.6
On Favorability	0.224 (p=0.015)	0.178 (p=0.053)	Coeffs. in Table E.7
On Qualification	0.205 (p=0.001)	0.16 (p=0.010)	Coeffs. in Table E.8
	Marginal Effect of Gender:		
	Change for Not Vets	Change for Vets	Location
On Leadership	-0.135 (p=0.194)	-0.015(p=0.889)	In-Text
			Figure 5
			Coeffs. in Table E.6
On Favorability	0.043 (p=0.643)	-0.002 (p=0.981)	Coeffs. in Table E.7
On Qualification	0.047 (p=0.453)	0.002 (p=0.980)	Coeffs. in Table E.8

Marginal effects reported in text for leadership. All results in this table are based on the estimates without controls in SI tables.

E.6 Results on Knowledge Measures

We also asked participants candidate knowledge about defense and the economy; coefficient estimates for those measures are in Table E.10.

Table E.10: Study 2: Effect of Treatment on Knowledge Perceptions (OLS)

	Model 1: Defense		Model 2: Defense		Model 1: Econ.		Model 2: Econ.	
	Coeff.	**(SE.)**	**Coeff**	**(SE)**	**Coeff**	**(SE)**	**Coeff**	**(SE)**
Woman Cand.	-0.055	0.108	-0.048	0.108	0.059	0.105	0.091	0.105
Vet	0.588**	0.107	0.581**	0.152	0.068	0.104	0.062	0.104
Woman × Vet	0.049	0.152	0.046	0.152	-0.023	0.148	-0.059	0.147
Ideology			-0.010	0.023			-0.044*	0.022
Woman Partic.			-0.042	0.082			-0.183*	0.079
White			-0.283**	0.097			-0.324**	0.093
Education			-0.041	0.030			-0.059*	0.028
Income			0.028	0.018			0.010	0.017
Independent			-0.493**	0.107			-0.600**	0.103
Intercept	4.129**	0.075	4.508**	0.186	4.211**	0.074	4.993	0.180
N	1,647		1,600		1,640		1,595	

*p<0.05, **p<0.01; OLS; outcomes: the politician's perceived defense and economic knowledge measures as outcome measure.

E.7 Study 2: Effect of Statements on Post-Treatment Positions

We present the results of the zero-sum funding measure as figures in the manuscript; coefficient estimates are shown in Table E.12, Table E.11 and Table E.13. The additional outcome measure included in the pre-registration is shown in Table E.15. We present the results without the controls in text; results with controls, excluding independents.

Table E.11: Effect of Treatment on Importance OLS, Zero Sum, Gender As Nuisance Factor (in-text)

	Controls		No Inds		In Text	
Variable	**Coeff.**	**(SE)**	**Coeff.**	**(SE)**	**Coeff**	**(SE)**
Vet	-0.047	(0.034)	-0.029	(0.037)	-0.046	(0.034)
Statement	0.081*	(0.034)	0.086*	(0.037)	0.082*	(0.034)
Vet × Statement	-0.062	(0.048)	-0.054	(0.052)	-0.056	(0.048)
Ideology	-0.009	(0.007)	—	—	—	—
White	-0.149**	(0.030)	—	—	—	—
Income	-0.006	(0.006)	—	—	—	—
Woman Partic.	-0.071**	(0.026)	—	—	—	—
Education	-0.024*	(0.009)	—	—	—	—
Independent	0.095**	(0.034)	—	—	—	—
Intercept	0.678**	(0.060)	0.365	(0.026)**	0.392**	(0.024)
N	1,609		1,399		1,649	

*p<0.05, **p<0.01, ***p<0.001; OLS; 0 not in line, 1 in line with statement. In-text column produced Figure 6

Table E.12: Effect of Treatment on Importance OLS, Zero Sum, Veteran Status as a Nuisance Factor (in-text)

	Controls		No Inds		In Text	
Variable	**Coeff.**	**(SE)**	**Coeff.**	**(SE)**	**Coeff**	**(SE)**
Woman Cand.	0.048	(0.033)	0.034	(0.037)	0.042	(0.034)
Statement	0.099**	(0.034)	0.119**	(0.037)	0.106*	(0.034)
Woman × State.	-0.097*	(0.048)	-0.119*	(0.052)	-0.104*	(0.048)
Ideology	-0.008	(0.007)	—	—	—	—
White	-0.150**	(0.030)	—	—	—	—
Income	-0.005	(0.006)	—	—	—	—
Woman Partic.	-0.067**	(0.026)	—	—	—	—
Education	-0.024*	(0.009)	—	—	—	—
Independent	0.097**	(0.034)	—	—	—	—
Intercept	0.627**	(0.059)	0.331**	(0.026)	0.349**	(0.024)
N	1,609		1,399		1,649	

*p<0.05, **p<0.01, ***p<0.001; OLS; 0 not in line, 1 in line with statement. In-text column produced Figure 7

Appendix F Study 2: Exploratory Analyses

F.1 Study 2: Effect of Statement on Trait Ratings

In our pre-registration we make predictions about the interaction of gender and veteran status on trait ratings, but do not make any predictions about the way these trait ratings will vary by the presence of the statement. However, we present the means in Table F.1.

Table E.13: Effect of Treatment on Importance OLS, Zero Sum (in-text)

	Controls		No Inds		In Text	
Variable	**Coeff.**	**(SE)**	**Coeff.**	**(SE)**	**Coeff**	**(SE)**
Woman Cand.	0.054	(0.048)	0.026	(0.052)	0.039	(0.048)
Vet	-0.042	(0.048)	-0.042	(0.052)	-0.050	(0.0478)
Woman × Vet	-0.011	(0.068)	0.023	(0.073)	0.007	(0.068)
Statement	0.129**	(0.048)	0.136**	(0.052)	0.127*	(0.048)
Woman × State.	-0.094	(0.068)	-0.099	(0.0735)	-0.089	(0.0678)
Vet × Statement	-0.058	(0.068)	-0.033	(0.073)	-0.041	(0.0679)
Wom. × Vet × State	-0.008	(0.096)	-0.045	(0.103)	-0.029	(0.096)
Ideology	-0.009	(0.007)	—	—	—	—
White	-0.147**	(0.030)	—	—	—	—
Income	-0.005	(0.006)	—	—	—	—
Woman Partic.	-0.070**	(0.026)	—	—	—	—
Education	-0.023*	(0.009)	—	—	—	—
Independent	0.096**	(0.034)	—	—	—	—
Intercept	0.649**	(0.064)	0.352**	(0.036)	0.373**	(0.0337)
N	1,609		1,399		1,649	

*p<0.05, **p<0.01, ***p<0.001; OLS; 0 not in line, 1 in line with statement. In-text column produced Figure 8

Table E.14: Effect of Treatment on Importance Logit, Zero Sum (additional results)

	Controls		No Inds		No Controls	
Variable	**Coeff.**	**(SE)**	**Coeff.**	**(SE)**	**Coeff**	**(SE)**
Woman Cand.	0.233	(0.210)	0.111	(0.223)	0.162	(0.202)
Vet	-0.193	(0.215)	-0.191	(0.229)	-0.219	(0.207)
Woman × Vet	-0.036	(0.299)	0.111	(0.318)	0.039	(0.290)
Statement	0.547*	(0.209)	0.563*	(0.219)	0.518*	(0.200)
Woman × State.	-0.399	(0.293)	-0.408	(0.310)	-0.365	(0.282)
Vet × Statement	-0.229	(0.297)	-0.112	(0.314)	-0.146	(0.287)
Wom. × Vet × State	-0.064	(0.420)	-0.218	(0.444)	-0.149	(0.406)
Ideology	-0.040	(0.032)	—	—	—	—
White	0.616**	(0.130)	—	—	—	—
Income	-0.023	(0.025)	—	—	—	—
Woman Partic.	-0.308*	(0.112)	—	—	—	—
Education	-0.105**	(0.042)	—	—	—	—
Independent	0.414*	(0.144)	—	—	—	—
Intercept	-0.663**	(0.280)	-0.609**	(0.158)	-0.518***	(0.143)
N	1,609		1,399		1,649	

*p<0.05, **p<0.01, ***p<0.001; Logit estimates; 0 not in line, 1 in line with statement.

Table E.15: Additional Importance Measure, OLS

	Controls Coef (SE)	**No Inds Coef (SE)**	**No Controls Coef (SE)**
Woman Cand.	-0.185 (0.146)	-0.224 (0.163)	-0.267 (0.148)
Vet	-0.280 (0.146)	-0.221 (0.163)	-0.334(0.148)*
Woman × Vet	0.159 (0.205)	0.198 (0.230)	0.250 (0.209)
Statement	-0.040 (0.146)	-0.034(0.163)	-0.083 (0.147)
Woman × Statement	0.141 (0.206)	0.215(0.231)	0.211 (0.209)
Vet × Statement	0.260 (0.206)	0.220 (0.230)	0.281 (0.209)
Woman × Vet × Statement	-0.235 (0.290)	-0.305 (0.326)	-0.265 (0.295)
Ideology	-0.189 (0.022)***	— (—)	— (—)
White	0.041 (0.092)	— (—)	— (—)
Income	0.003 (0.017)	— (—)	— (—)
Woman Partic.	-0.141 (0.078)	— (—)	— (—)
Education	-0.053 (0.029)	— (—)	— (—)
Independent	0.388 (0.102***	— (—)	— (—)
Constant	4.710** (0.194)	3.689(0.114)**	3.824(0.104)**
N	1,612	1,402	1,652

*p<0.05, **p<0.01; OLS, using an alternative measure of spending importance, re-coded to match the direction of the statement participants received.

F.1.1 Effect of Statement on Leadership: Democrats Only

Next we conduct a comparison that addresses the role of the statement, for *Democratic* candidates. This is because the Democratic candidate makes a statement that could be interpreted as hawkish. We did not pre-register this idea because we did not intend for this to be an expressly hawkish signal. This point, however, was raised during the review process. We present our results in Table F.2. We do not find that the statement increases perceptions of leadership (and the effect of the statement is not significant).

Table E.16: Means by Group

	Defense Spending *(Follow Statement Direction)*	**Defense - Zero Sum** *(Follow Statement Direction)*
Woman, Not a Vet, No Statement	3.56 (0.105)	0.41 (0.035)
Woman, Vet, No Statement	3.47 (0.104)	0.37 (0.033)
Woman, Not a Vet, Statement	3.68 (0.104)	0.45 (0.035)
Woman, Vet, Statement	3.62 (0.099)	0.34 (0.033)
Man, Not a Vet, No Statement	3.82 (0.109)	0.37 (0.034)
Man, Vet, No Statement	3.49 (0.108)	0.32 (0.033)
Man, Not a Vet, Statement	3.74 (0.103)	0.50 (0.035)
Man, Vet, Statement	3.69 (0.103)	0.41(0.034)
N	1,652	1,649

Only means and standard errors presented.

F.2 Study 2: Results by Subgroups

We next consider results by subgroups(see Table F.3).

Gender: Substantive results for participants who are women (changes in probability due to statement): (1) women who are not veterans: + 0.015 (p=0.815); (2) women who are veterans: -0.028 (0.675); (3) men who are not veterans: 0.168 (p=0.010); (4) men who are veterans: +0.055 (p=0.423).

Substantive results for participants who are men (changes in probability due to statement): (1) women who are not veterans: +0.064 (p=0.361); (2) women who are veterans: -0.025 (p=0.713); (3) men who are not veterans:+ 0.075 (p=0.285); (4) men who are veterans: -0.107 (p=0.118).

Party Substantive results for Democratic participants (changes in probability due to statement, OLS): (1) women who are not veterans: +0.10 (p=0.167); (2) women who are veterans: -0.05 (0.511); (3) men who are not veterans +0.11 (p=0.123); (4) men who are veterans: by +0.11 (p=0.126).

Substantive results for Republican participants (changes in probability due to statement, OLS): (1) women who are not veterans: -0.036 (p=0.633); (2) women who are veterans: -0.029 (p=0.693); (3) men who are not veterans: +0.168 (p=0.031); (4) men who are veterans: +0.093 (p=0.211).

Ideology Substantive results for conservative participants (changes in probability, due to statement): (1) women who are not veterans: + 0.030 (p=0.734); (2) women who are

Table F.1: Means by Group

	Leadership	**Favorability**	**Knowledge Defense**	**Knowledge Economy**	**Qualified**
Woman, Not a Vet, No Statement	4.43 (0.103)	4.39 (0.089)	3.99 (0.117)	4.18 (0.115)	3.40 (0.066)
Woman, Vet, No Statement	4.74 (0.098)	4.59 (0.085)	4.98 (0.100)	4.44 (0.094)	3.59 (0.058)
Woman, Not a Vet, Statement	4.19 (0.109)	4.03 (0.102)	4.15 (0.115)	4.36 (0.114)	3.23 (0.066)
Woman, Vet, Statement	4.44 (0.106)	4.18 (0.095)	4.45 (0.108)	4.20 (0.102)	3.36 (0.058)
Man, Not a Vet, No Statement	4.52 (0.097)	4.31 (0.080)	4.15 (0.101)	4.26 (0.100)	3.33 (0.056)
Man, Vet, No Statement	4.80 (0.098)	4.58 (0.081)	4.82 (0.104)	4.12 (0.098)	3.57 (0.059)
Man, Not a Vet, Statement	4.37 (0.111)	4.01 (0.095)	4.11 (0.102)	4.16 (0.104)	3.20 (0.065)
Man, Vet, Statement	4.41 (0.105)	4.19(0.103)	4.61 (0.109)	4.44 (0.106)	3.37(0.068)
N	1,643	1,659	1,660	1,647	1,640

Only means and standard errors presented.

Table F.2: Study 2, Leadership Evaluations, Democratic Candidate Only

	Statement	**No Statement**	**Difference**
Woman, Not a Veteran	4.29	4.45	p=0.457
Woman, Veteran	4.40	4.86	p=0.033
Man, Not a Veteran	4.44	4.37	p=0.727
Man, Veteran	4.28	4.76	p=0.030

Analyses are t-tests; N=763

veterans: -0.002 (0.982); (3) men who are not veterans:+ 0.117 (p=0.202); (4) men who are veterans: +0.105 (p=0.227).

Substantive results for liberal participants (changes in probability changes due to statement): (1) women who are not veterans:+0.076 (p=0.314); (2) women who are veterans: -0.081 (p=0.309); (3) men who are not veterans: +0.092 (p=0.237); (4) men who are veterans: +0.192 (p=0.017).

Education Substantive results for high education participants, BA and above (changes in probability due to statement): (1) women who are not veterans: +0.066 (p=0.404); (2) women who are veterans: -0.12 (0.132); (3) men who are not veterans: +0.086 (p=0.291); (4) men who are veterans: +0.123 (p=0.144).

Substantive results for low education participants (changes in probability due to statement): (1) women who are not veterans: +0.012 (p=0.840); (2) women who are veterans: +0.025 (p=0.674); (3) men who are not veterans: 0.148 (p=0.013); (4) men who are veterans: +0.062 (p=0.298).

Table F.3: Zero-Sum Measure by Groups, OLS Results

	Women	Men	Dem	Rep	Cons.	Lib.	H.E.	L.E.
	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
Woman	0.102	-0.044	-0.008	0.065	0.010	0.017	-0.021	0.078
	(0.066)	(0.070)	(0.071)	(0.075)	(0.089)	(0.076)	(0.077)	(0.061)
Vet	0.018	-0.143*	-0.003	-0.089	-0.062	-0.120	-0.102	-0.021
	(0.066)	(0.042)	(0.072)	(0.074)	(0.086)	(0.080)	(0.082)	(0.059)
Woman × Vet	-0.088	0.114	0.015	0.033	0.112	0.013	0.101	-0.053
	(0.095)	(0.098)	(0.101)	(0.105)	(0.123)	(0.110)	(0.114)	(0.085)
Statement	0.168*	0.080	0.108	0.168*	0.117	0.092	0.086	0.148*
	(0.065)	(0.071)	(0.070)	(0.078)	(0.091)	(0.078)	(0.081)	(0.059)
Woman × State.	-0.153	-0.016	-0.009	-0.204	-0.086	-0.016	-0.020	-0.135
	(0.093)	(0.100)	(0.100)	(0.108)	(0.128)	(0.109)	(0.113)	(0.085)
Vet × State.	-0.113	0.026	0.002	-0.075	-0.012	0.100	0.037	-0.086
	(0.094)	(0.098)	(0.100)	(0.108)	(0.126)	(0.112)	(0.117)	(0.084)
Wom. × Vet × State.	0.070	-0.119	-0.148	0.082	-0.020	-0.257	-0.223	0.099
	(0.133)	(0.139)	(0.142)	(0.150)	(0.174)	(0.157)	(0.162)	(0.120)
Constant	0.288**	0.479**	0.383**	0.317**	0.354**	0.373**	0.386**	0.367**
	(0.045)	(0.050)	(0.051)	(0.052)	(0.061)	(0.055)	(0.057)	(0.042)
N	820	814	770	629	515	589	569	1078

*p<0.05, **p<0.01, OLS; dependent variable: 0 not in line with statement, 1 in line.

Appendix G Study 2: Pre-Registration and Deviations

The full version of our pre-registration is here http://aspredicted.org/blind.php?x=mw6ir9 and is also in the supplementary appendix (last page).

We note several deviations from our pre-registered analyses.

First, we did not consider military spending as an outcome variable as we ultimately did not measure this variable due to the length of the survey, as this variable was largely exploratory and we did not have any *a priori* expectations, this measure was cut.

Second, we also did not consider economic attitudes as a covariate, again due to the length of the survey and the fact that we did not have *a priori* hypotheses.

In checks, we also control for whether someone is an independent (as those participants were randomly assigned to a party) and ideology, in addition to basic demographics.

In our pre-registration, we note that we would use t-tests. Although we present means and comparisons between means, we use OLS models that only include the treatments (and some with controls as checks – as pre-registered). This was due to suggestions from readers on the first drafts of the manuscript. We also pre-registered the possibility of using a difference-in-difference analysis, which was determined not to be appropriate in our case – we do report the answer to our research question, however (no gender differences).

In our pre-registration we also pre-registered conditioning on a post-treatment check. We report the results of this check, but do not condition on it, as after our pre-registration was filed, numerous articles have shown that this would introduce bias.

Readers of the manuscript were interested in the role of gender (across veteran status) in response to the statement. We present this analysis – which we specify is exploratory.

During the review process we were asked a number of questions about the effect of the statement on evaluations of leadership – we did not have any pre-registered expectations to this point. We include the patterns in the SI.

During the review process, reviewers suggested additional literature and scholarship to include in our discussion of leadership and authority. The suggestions helped us clarify the distinction between leadership and authority; we note, however, that some of the suggested scholarship was published after we pre-registered and fielded the study so it could not have informed our a priori arguments.

Gender and Veteran Status (#16193)

Created: 11/09/2018 02:44 PM (PT)
Shared: 01/12/2019 09:33 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Study tests the following hypotheses: Hypotheses Within Gender: 1. Candidates (across both genders) who are combat veterans are perceived as more likely to be strong leaders relative to candidates who are not combat veterans. 2. Candidates (across both genders) who are combat veterans will be more successful at changing defense spending preferences than those who are not combat veterans Hypotheses Across Gender: 1. Men who are combat veterans will be perceived as stronger leaders than women candidates who are combat veterans 2. Men who are combat veterans will be more successful at changing defense spending preferences than women candidates who are combat veterans RQ1: Is there a difference in the effect of being described as a combat veteran for men and women?

3) Describe the key dependent variable(s) specifying how they will be measured.
1. Favorability of candidate, 2. Perceived leadership and knowledgeability on topics, 4. Perceived importance of defense spending relative to other spending, 5. Zero-sum defense spending question.

4) How many and which conditions will participants be assigned to?
Participants are randomly assigned across 8 groups, comprised of the following factors/levels: Factor 1: Statement on defense or no statement (two levels) Factor 2: Combat Veteran or not (two levels) Factor 3: Woman or Man candidate (two levels)

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will conduct analyses along the hypotheses comparisons: within gender comparison will consider groups within gender, across gender comparison will consider candidates within type, but across gender. We will rely on (a) t-tests and (b) models with basic controls measured pre-treatment (basic demographics). Research questions require us to conduct a difference-in-difference analyses.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Since this will be conducted via RNSSI, there is the possibility that some participants take an unusually long time. We will first look at the times it has taken participants to complete the study, excluding those who are outliers -- if such outliers exist. We will also test the analyses focusing only on people who pass a manipulation check at the end of the study (which asks them if they recall the gender of the candidate in their treatment)

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
1600 participants have been requested from RN SSI.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will be conducing a number of exploratory analyses. First, research suggests that Democrats and Republicans and Liberals and Conservatives will behave in different ways. Therefore, we will consider the patterns for the parties/ideological groups separately. Second, we will consider whether men and women differ, though we do not have any precise hypotheses. Third, we will consider whether people's economic predispositions affect their responses to the treatment. Fourth, we will consider whether education affects response to treatment as education is often a proxy for tolerance toward other groups. Finally, in the past we have conducted studies that consider who can influence attitudes toward military spending. We will consider these patterns here as well, using military spending as a dependent variable. We also want to note that this is an update to a prior pre-registration for a version of this study that was never run. We revised the study after pre-registration but before fielding to remove a condition, hence the new pre-registration.

Appendix H Human Subjects Approval