

Supplementary Materials for: The Gendered Peace Premium

Christopher W. Blair and Joshua A. Schwartz

May 30, 2023

Contents

A-1 Study 1	A-2
A-1.1 Results with a 7-Point Scale of Disapproval	A-2
A-1.2 Results with Respondents Who Passed Manipulation Check	A-3
A-1.3 Results with Covariates	A-4
A-1.4 Mediation Analyses	A-5
A-1.5 Heterogeneity in the Gendered Peace Premium	A-6
A-1.6 Probing Co-Partisan and Out-Partisan Dynamics	A-7
A-1.7 Do Leader Sex and Partisanship Interact?	A-8
A-1.8 Other Pre-Registered Heterogeneous Effects	A-9
A-2 Study 2	A-10
A-2.1 Manipulating Leader Sex, Party, and Disposition	A-10
A-2.2 Results with a 7-Point Scale of Disapproval	A-11
A-2.3 Results with Respondents Who Passed Manipulation Check	A-12
A-2.4 Results with Covariates	A-13
A-2.5 Do Leader Gender, Partisanship, and Disposition Interact?	A-14
A-2.6 Mediation Analysis	A-15
A-2.7 Heterogeneity in the Gendered Peace Premium	A-17
A-2.8 Other Pre-Registered Heterogeneous Effects	A-18
A-3 References for Supplementary Materials	A-19

A-1 Study 1

A-1.1 Results with a 7-Point Scale of Disapproval

Table A.1 shows that our results are robust to using the full 7-point scale of disapproval, rather than the binary measure of disapproval as in the main text. There is a significant gendered peace premium, but no distinguishable partisan premium.

Table A.1: Study 1 — 7-Point Scale of Disapproval

	Disapproval (7-Point Scale)		
	Conciliatory Policy	Status Quo Policy	Effect of Conciliation (7-Point Scale)
Male President	3.827*** (3.608, 4.046)	2.788*** (2.604, 2.973)	1.039*** (0.753, 1.326)
Female President	4.212*** (3.987, 4.436)	2.623*** (2.456, 2.790)	1.589*** (1.309, 1.869)
Gendered Peace Premium			0.550*** (0.150, 0.950)
Republican President	3.941*** (3.729, 4.154)	2.759*** (2.576, 2.942)	1.183*** (0.902, 1.463)
Democratic President	4.100*** (3.867, 4.333)	2.650*** (2.481, 2.820)	1.450*** (1.161, 1.738)
Partisan Peace Premium			0.267 (-0.135, 0.669)

Note: *p<0.10, **p< 0.05, ***p<0.01. 95% confidence intervals are in parentheses.

A-1.2 Results with Respondents Who Passed Manipulation Check

Table A.2 also shows that our main results are robust to excluding respondents that failed relevant factual manipulation check questions and thus were not paying close attention to our vignette. Unsurprisingly, the female peace premium is slightly greater among respondents that were paying attention to our treatments (14.5 percentage points) than among all respondents (11.6 percentage points).

Table A.2: Study 1 — Respondents Who Passed Factual Manipulation Check

	Disapproval (% Points)		
	Conciliatory Policy	Status Quo Policy	Effect of Conciliation (% Points)
Male President	39.130*** (30.154, 48.107)	9.091*** (3.936, 14.246)	30.040*** (19.689, 40.390)
Female President	51.063*** (42.761, 59.367)	6.569*** (2.395, 10.744)	44.495*** (35.201, 53.788)
Gendered Peace Premium			14.455** (0.544, 28.366)
Republican President	41.060*** (33.174, 48.946)	9.091*** (4.682, 13.500)	31.969*** (22.934, 41.003)
Democratic President	47.273*** (39.617, 54.929)	8.176*** (3.896, 12.457)	39.097*** (30.325, 47.868)
Partisan Peace Premium			7.128 (-5.464, 19.720)

Note: *p<0.10, **p< 0.05, ***p<0.01. 95% confidence intervals are in parentheses.

A-1.3 Results with Covariates

Table A.3 illustrates that our results hold in a regression context controlling for other covariates. This includes respondent perceptions of whether the president was white or not, which helps account for the possibility of a lack of information equivalence across experimental conditions where manipulating one factor—like gender—leads respondents to update their beliefs about other relevant but not experimentally manipulated or controlled dimensions—like leader race (Dafoe, Zhang, and Caughey, 2018). The quantities of interest are not the regression coefficients themselves, but the peace premia calculations at the bottom of the table. As in the main text, there is a statistically significant gendered peace premium, but not a significant partisan premium.

Table A.3: Study 1 — Results with Covariates

	Gendered Peace Premium		Partisan Peace Premium	
	(1) Disapproval (% Points)	(2) Disapproval (7-Point Scale)	(3) Disapproval (% Points)	(4) Disapproval (7-Point Scale)
Male x Status Quo	-0.404*** (-0.615, -0.193)	1.753*** (0.909, 2.597)		
Male x Conciliatory	-0.175* (-0.381, 0.032)	2.782*** (1.955, 3.608)		
Female x Status Quo	-0.440*** (-0.646, -0.234)	1.567*** (0.734, 2.399)		
Female x Conciliatory	-0.071 (-0.277, 0.134)	3.174*** (2.348, 4.001)		
Republican x Status Quo			-0.418*** (-0.630, -0.206)	1.694*** (0.848, 2.540)
Republican x Conciliatory			-0.147* (-0.355, 0.060)	2.896*** (2.068, 3.724)
Democratic x Status Quo			-0.432*** (-0.643, -0.222)	1.571*** (0.723, 2.419)
Democratic x Conciliatory			-0.103 (-0.311, 0.106)	3.015*** (2.174, 3.855)
Democratic President	0.016 (-0.037, 0.069)	0.001 (-0.208, 0.209)		
Female President			0.032 (-0.022, 0.087)	0.099 (-0.114, 0.312)
Respondent-Level Covariates	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓
Observations	813	813	813	813
Peace Premium:	13.901** (3.235, 24.568)	0.579*** (0.165, 0.993)	5.879 (-4.827, 16.586)	0.242 (-0.176, 0.661)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-1.4 Mediation Analyses

A formal causal mediation analysis shows that policy credibility and competence are significant mechanisms explaining the gendered peace premium (Imai, Keele, and Tingley, 2010; Hicks and Tingley, 2011).

Table A.4: Study 1 — Average Causal Mediation Effect of Policy Credibility

	Male President	Female President
Average Mediation Effect	0.858 (0.589, 1.138)	1.307 (1.022, 1.608)
Average Direct Effect	0.175 (-0.005, 0.359)	0.310 (0.132, 0.493)
Total Effect	1.033 (0.935, 1.139)	1.618 (1.495, 1.747)
% of Total Effect Mediated	83.183% (75.344, 91.760)	80.856% (74.878, 87.460)
Respondent-Level Covariates	✓	✓
President's Perceived Race	✓	✓

Note: 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

Table A.5: Study 1 — Average Causal Mediation Effect of Competence

	Male President	Female President
Average Mediation Effect	0.529 (0.317, 0.758)	0.837 (0.598, 1.094)
Average Direct Effect	0.510 (0.272, 0.754)	0.780 (0.555, 1.011)
Total Effect	1.039 (0.964, 1.106)	1.618 (1.532, 1.700)
% of Total Effect Mediated	50.786% (47.845, 54.867)	51.693% (49.242, 54.645)
Respondent-Level Covariates	✓	✓
President's Perceived Race	✓	✓

Note: 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-1.5 Heterogeneity in the Gendered Peace Premium

In Table A.6 we test a number of heterogeneous effect hypotheses we had outlined in our pre-analysis plan. We find no evidence of heterogeneous effects, *except* with respect to partisanship (column 1). Republican respondents are significantly more likely to punish female leaders for pursuing peace than non-Republican respondents. In results omitted for space but available upon request, we show that this effect also holds when using a binary measure of disapproval, and if we drop non-partisan (i.e., non party-affiliated) respondents from the analysis. No similar relationship holds when analyzing the partisan peace premium.

Table A.6: Study 1 — Heterogeneity in the Gendered Peace Premium

Heterogeneity by Respondent:	Disapproval (7-Point Scale)						
	(1) Partisanship	(2) Hostile Sexism	(3) Benevolent Sexism	(4) Second-Order Sexism	(5) Militant Assertiveness	(6) Education	(7) Gender
Republican Respondent	1.287*** (0.537, 2.036)						
Non-Republican Respondent	0.247 (-0.231, 0.724)						
Difference in Premia	1.040** (0.138, 1.942)						
High Sexism Respondent		0.568 (-0.116, 1.251)	0.505 (-0.170, 1.179)	0.366 (-0.410, 1.142)			
Low Sexism Respondent		0.761* (-0.070, 1.591)	0.621 (-0.230, 1.471)	0.855*** (0.217, 1.493)			
Difference in Premia		-0.193 (-1.276, 0.890)	-0.116 (-1.207, 0.976)	-0.489 (-1.499, 0.521)			
Hawkish Respondent					0.434 (-0.282, 1.150)		
Dovish Respondent					0.381 (-0.336, 1.097)		
Difference in Premia					0.053 (-0.965, 1.071)		
High Education Respondent						0.708** (0.054, 1.362)	
Low Education Respondent						0.343 (-0.439, 1.125)	
Difference in Premia						0.365 (-0.658, 1.388)	
Female Respondent							0.465 (-0.109, 1.039)
Male Respondent							0.709** (0.106, 1.312)
Difference in Premia							-0.244 (-1.080, 0.592)
Respondent-Level Covariates	✓	✓	✓	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓	✓	✓	✓
Observations	813	556	525	597	526	562	813

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-1.6 Probing Co-Partisan and Out-Partisan Dynamics

Do respondents impose a more severe penalty on out-partisan women presidents? We test this dynamic in Table A.7. We define co- and out-partisan presidents by reference to the president’s partisan identification and respondents’ partisan identity. In Study 1, we do not find a distinguishable difference in the gendered premium between co-partisan and out-partisan executives. Rather, a discernable penalty is imposed on women for conciliation irrespective of their partisan relationship to respondents.

Table A.7: Study 1 — Co-Partisans vs. Out-Partisans

	Disapproval (7-Point Scale)		
	(1) Sample of Co-Partisans	(2) Sample of Out-Partisans	(3) Sample of All Respondents
Male x Status Quo	0.375 (-1.214, 1.965)	1.999** (0.468, 3.531)	
Male x Conciliatory	1.434* (-0.185, 3.053)	3.084*** (1.560, 4.608)	
Female x Status Quo	0.008 (-1.589, 1.606)	1.904** (0.408, 3.400)	
Female x Conciliatory	1.870** (0.254, 3.487)	3.707*** (2.206, 5.209)	
Male x Status Quo x Co-Partisan			1.445*** (0.586, 2.303)
Male x Conciliatory x Co-Partisan			2.415*** (1.550, 3.280)
Female x Status Quo x Co-Partisan			1.205*** (0.354, 2.058)
Female x Conciliatory x Co-Partisan			2.711*** (1.856, 3.566)
Male x Status Quo x Out-Partisan			0.357 (-0.085, 0.798)
Male x Conciliatory x Out-Partisan			0.510** (0.026, 0.995)
Female x Status Quo x Out-Partisan			0.494** (0.095, 0.893)
Female x Conciliatory x Out-Partisan			0.772*** (0.326, 1.218)
Democratic President	0.573 (-0.301, 1.447)	0.568 (-0.349, 1.484)	0.068 (-0.143, 0.280)
Respondent-Level Covariates	✓	✓	✓
President’s Perceived Race	✓	✓	✓
Observations	331	312	813
Co-Partisans	0.804** (0.160, 1.447)		0.659* (-0.050, 1.368)
Out-Partisans		0.718* (-0.0004, 1.437)	0.535** (0.039, 1.030)
Difference in Premia			0.124 (-0.739, 0.988)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President’s perceived race is each respondent’s belief about the race of the president.

A-1.7 Do Leader Sex and Partisanship Interact?

By varying gender and partisanship, some treatment vignettes offer potentially contradictory signals about a leader’s disposition towards conciliation. For instance, a Republican woman president may be perceived as more hawkish than a Democratic woman president. It is not clear how respondents will interpret conflicting signals from leader gender and leader partisanship. We study this dynamic in a series of exploratory analyses in Table A.8. Specifically, we calculate gendered peace premia for Democratic and Republican presidents in a regression framework. There are no distinguishable differences in premia across these conditions.

Table A.8: Study 1 — Gendered Peace Premia for Democratic vs. Republican Presidents

	Sample: All Respondents		Sample: Passed Manipulation Check	
	Disapproval (% Points)	Disapproval (7-Point Scale)	Disapproval (% Points)	Disapproval (7-Point Scale)
Democratic Presidents	8.870 (-5.957, 23.696)	0.659** (0.084, 1.233)	14.247 (-4.873, 33.367)	0.719* (-0.027, 1.465)
Republican Presidents	14.533* (-0.304, 29.370)	0.448 (-0.112, 1.008)	15.963 (-4.281, 36.207)	0.395 (-0.370, 1.160)
Difference in Premia	-5.664 (-26.639, 15.312)	0.211 (-0.592, 1.013)	-1.716 (-29.562, 26.130)	0.324 (-0.744, 1.392)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. The Republican estimate is marginally imprecise in columns 2 ($p = 0.117$) and 3 ($p = 0.122$).

A-1.8 Other Pre-Registered Heterogeneous Effects

In addition to the heterogeneous effect analyses in Table A.6, we also pre-registered a number of additional expectations. First, we anticipated that hostile sexists would be more disapproving of women leaders in general. In column 1 we do not find evidence in support. Second, following from Mattes and Weeks (2019), we expected hawkish respondents to be more disapproving of conciliation irrespective of leader sex. Results in column 2 show hawkish respondents are more likely to disapprove of leaders pursuing conciliation. Because Republicans also tend to be more hawkish (Kertzer and Brutger, 2016), we also expected Republicans to be more disapproving of conciliation irrespective of leader sex. In column 3 we find that this is the case. Finally, column 4 considers whether female respondents are less disapproving of women presidents irrespective of their policies (conciliatory or status quo). We find no evidence for this hypothesis.

Table A.9: Study 1 — Other Pre-Registered Heterogeneous Effects

	Disapproval (7-Point Scale)			
	(1)	(2)	(3)	(4)
Female President × Hostile Sexism	0.127 (-0.107, 0.361)			
Conciliatory Policy × Militant Assertiveness		1.153*** (0.910, 1.397)		
Conciliatory Policy × Republican Respondent			0.325*** (0.220, 0.430)	
Female President × Female Respondent				0.077 (-0.343, 0.497)
Female President	-0.293 (-1.029, 0.442)	0.123 (-0.077, 0.322)	0.118 (-0.090, 0.327)	0.057 (-0.248, 0.362)
Conciliatory Policy	1.327*** (1.118, 1.536)	-2.354*** (-3.146, -1.562)	0.122 (-0.307, 0.551)	1.325*** (1.115, 1.534)
Hostile Sexism	0.048 (-0.133, 0.229)			
Militant Assertiveness		-0.543*** (-0.724, -0.361)		
Republican Respondent			-0.061 (-0.136, 0.014)	
Female Respondent				-0.004 (-0.311, 0.303)
Democratic President	-0.006 (-0.214, 0.203)	0.021 (-0.177, 0.218)	0.019 (-0.185, 0.223)	-0.005 (-0.214, 0.204)
Respondent-Level Covariates	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓
Observations	813	813	813	813

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-2 Study 2

A-2.1 Manipulating Leader Sex, Party, and Disposition

As noted in the manuscript, in Study 2 we experimentally vary a leader’s sex, partisanship, policy choice, and foreign policy disposition. This marks a departure from Study 1, which only varies a leader’s sex, partisanship, and policy choice. We expect gender stereotypes to influence evaluations of women leaders’ peace proposals by shaping assessments of women leaders’ foreign policy orientations. In Study 1 we argue and find evidence that women face a higher penalty than men for pursuing peace versus the status quo because women are (gender-stereotypically) perceived as more dovish. By manipulating a leader’s disposition in Study 2 we block the proposed mechanism by which leader sex shapes attitudes in Study 1. In addition to this overarching reason, two more minor considerations are also relevant for our decision not to directly vary leader disposition in Study 1.

First, a strong hawk–dove treatment may be unrealistic, reducing the generalizability of a scenario in which it is varied. For one, a strong dispositional prime implies that foreign policy reputation is objective, and that leaders develop clear hawk–dove positions. In the real world, foreign policy reputation is subjective and ambiguous. After all, leaders often adopt hawkish *and* dovish positions on comparable foreign policy issues—if they adopt substantive positions at all (Ziv, 2014). For example, President Obama withdrew from Iraq and signed a nuclear deal with Iran, but also escalated the war in Afghanistan and expanded drone use. President Trump signed a peace deal with the Taliban and sought rapprochement with North Korea, but also withdrew from the Iran nuclear deal, assassinated Qasem Soleimani, and adopted hawkish policies towards China. Contrary actions like this blur the distinction between hawk–dove categories, making it more difficult for leaders to establish consistent dispositional reputations (Renshon, Dafoe, and Huth, 2018).

Second, a strong hawk–dove treatment assumes the public pays close attention to a leader’s past foreign policy positions. Although public opinion is central to foreign policy decisionmaking in democracies (Tomz, Weeks, and Yarhi-Milo, 2020), the mass public is often poorly informed about foreign policy in general. This means members of the public frequently lack a clear sense of leaders’ past foreign policy positions and any reputations they have established among elites (Guisinger and Saunders, 2017; Kertzer, Brooks, and Brooks, 2021). By contrast, there is broad awareness of and typically little to no uncertainty about a leader’s sex and gender identity. The issue of dispositional ambiguity is emblemized by debates over Hillary Clinton’s foreign policy orientation. Although some journalists and officials argued she was a hawk, others argued she was actually dovish and preferred diplomatic over military solutions.⁰ More objective measures of foreign policy orientation often suggest that differences between politicians are relatively small. For instance, Bendix and Jeong (2019) calculate foreign policy scores for each member of Congress from 1993 to 2016, with higher scores indicating greater hawkishness. Hillary Clinton’s average score in the Senate was -1.502, while Barack Obama’s was -1.535 and John Kerry’s was -1.419. Despite arguments that Clinton was more hawk-

⁰See articles in [CNN](#), [NPR](#), the [Wall Street Journal](#), and [Vox](#).

ish than Obama or Kerry, this measure shows scant difference. Similarly, a YouGov poll conducted in 2014 found about a third of Americans believed Clinton was a hawk, a third believed she was a dove, and a third were not sure.⁰ The fact that a leader as prominent as Clinton failed to establish a clear dispositional reputation in the public’s eye suggests that sex, partisanship, and other cues may be more salient heuristics for the public in complex real-world scenarios (Trager and Vavreck, 2011; Guisinger and Saunders, 2017; Kertzer, Brooks, and Brooks, 2021).

A-2.2 Results with a 7-Point Scale of Disapproval

Table A.10 shows that our results are robust to using the full 7-point scale of disapproval, rather than the binary measure of disapproval as in the main text. There is a significant dispositional peace premium, but no distinguishable gendered or partisan premia.

Table A.10: Study 2 — 7-Point Scale of Disapproval

	Disapproval (7-Point Scale)		
	Conciliatory Policy	Status Quo Policy	Effect of Conciliation (7-Point Scale)
Male President	3.434*** (3.298, 3.571)	2.793*** (2.669, 2.916)	0.641*** (0.457, 0.826)
Female President	3.533*** (3.390, 3.677)	2.840*** (2.709, 2.970)	0.694*** (0.500, 0.887)
Gendered Peace Premium			0.052 (-0.215, 0.319)
Republican President	3.379*** (3.245, 3.512)	2.791*** (2.667, 2.915)	0.588*** (0.406, 0.770)
Democratic President	3.588*** (3.442, 3.733)	2.843*** (2.713, 2.974)	0.745*** (0.549, 0.940)
Partisan Peace Premium			0.157 (-0.110, 0.424)
Hawkish President	3.562*** (3.425, 3.700)	3.118*** (2.977, 3.258)	0.445*** (0.248, 0.642)
Dovish President	3.406*** (3.264, 3.548)	2.514*** (2.408, 2.620)	0.892*** (0.714, 1.069)
Dispositional Peace Premium			0.447*** (0.182, 0.712)

Note: *p<0.10, **p< 0.05, ***p<0.01. 95% confidence intervals are in parentheses.

⁰A Pew poll in 2008 found similar diversity in opinion about Clinton. A YouGov poll about Obama in 2014 also found highly mixed views about the president’s disposition.

A-2.3 Results with Respondents Who Passed Manipulation Check

Table A.11 also shows that our main results are robust to excluding respondents that failed relevant factual manipulation check questions and thus were not paying close attention to our vignette. Unsurprisingly, the hawk’s advantage is greater among respondents that were paying attention to our treatments (18.6 percentage points) than among all respondents (12.8 percentage points). Interestingly, this analysis also yields distinguishable evidence that a partisan peace premium exists: the net utility of rapprochement is 7.6 percentage points less for Democrats than Republicans. However, this result should be interpreted with caution since dropping respondents based on a post-treatment variable can lead to bias (Aronow et al. 2020). Moreover, this finding does not hold when using the full 7-point measure of approval and also excluding respondents that failed the factual manipulation check.

Table A.11: Study 2 — Respondents Who Passed Factual Manipulation Check

	Disapproval (% Points)		
	Conciliatory Policy	Status Quo Policy	Effect of Conciliation (% Points)
Male President	31.035*** (25.406, 36.663)	12.406*** (8.434, 16.379)	18.629*** (11.740, 25.517)
Female President	35.474*** (30.274, 40.674)	14.706*** (10.727, 18.685)	20.768*** (14.220, 27.316)
Gendered Peace Premium			2.140 (-7.365, 11.644)
Republican President	29.058*** (24.495, 33.620)	13.384*** (10.023, 16.744)	15.674*** (10.007, 21.340)
Democratic President	35.990*** (31.210, 40.769)	12.707*** (9.269, 16.145)	23.283*** (17.395, 29.170)
Partisan Peace Premium			7.609* (-0.563, 15.781)
Hawkish President	33.818*** (28.213, 39.424)	24.375*** (19.659, 29.091)	9.443** (2.118, 16.769)
Dovish President	32.615*** (27.832, 37.397)	4.587*** (2.314, 6.860)	28.027*** (22.732, 33.322)
Dispositional Peace Premium			18.584*** (9.546, 27.623)

Note: *p<0.10, **p< 0.05, ***p<0.01. 95% confidence intervals are in parentheses.

A-2.4 Results with Covariates

Table A.12 illustrates that our results hold in a regression context controlling for other covariates. The quantities of interest are not the regression coefficients themselves, but the peace premia calculations at the bottom of the table. As in the main text, there is a statistically significant dispositional peace premium, but no distinguishable gendered or partisan premia.

Table A.12: Study 2 — Results with Covariates

	Gendered Peace Premium		Partisan Peace Premium		Dispositional Peace Premium	
	(1) Disapproval (% Points)	(2) Disapproval (7-Point Scale)	(3) Disapproval (% Points)	(4) Disapproval (7-Point Scale)	Disapproval (% Points)	Disapproval (7-Point Scale)
Male x Status Quo	0.240*** (0.090, 0.390)	3.873*** (3.265, 4.481)				
Male x Conciliatory	0.386*** (0.242, 0.531)	4.549*** (3.957, 5.142)				
Female x Status Quo	0.267*** (0.118, 0.417)	3.946*** (3.338, 4.554)				
Female x Conciliatory	0.435*** (0.286, 0.583)	4.645*** (4.040, 5.251)				
Republican x Status Quo			0.246*** (0.098, 0.395)	3.898*** (3.297, 4.500)		
Republican x Conciliatory			0.377*** (0.232, 0.523)	4.521*** (3.923, 5.119)		
Democratic x Status Quo			0.245*** (0.098, 0.392)	3.980*** (3.376, 4.585)		
Democratic x Conciliatory			0.428*** (0.283, 0.572)	4.733*** (4.142, 5.324)		
Hawkish x Status Quo					0.274*** (0.125, 0.422)	4.023*** (3.421, 4.626)
Hawkish x Conciliatory					0.360*** (0.216, 0.503)	4.434*** (3.845, 5.023)
Dovish x Status Quo					0.125* (-0.014, 0.265)	3.315*** (2.737, 3.894)
Dovish x Conciliatory					0.353*** (0.212, 0.494)	4.281*** (3.701, 4.861)
Democratic President	0.025 (-0.010, 0.059)	0.147** (0.010, 0.285)			0.025 (-0.009, 0.060)	0.150** (0.013, 0.287)
Female President			0.038** (0.004, 0.072)	0.085 (-0.052, 0.221)	0.038** (0.004, 0.072)	0.085 (-0.051, 0.221)
Dovish President	-0.077*** (-0.111, -0.042)	-0.428*** (-0.565, -0.291)	-0.076*** (-0.111, -0.042)	-0.427*** (-0.564, -0.290)		
Respondent-Level Covariates	✓	✓	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓	✓	✓
Observations	1970	1970	1970	1970	1970	1970
Peace Premium:	2.098 (-4.793, 8.989)	0.023 (-0.250, 0.296)	5.160 (-1.695, 12.016)	0.130 (-0.143, 0.402)	14.172*** (7.281, 21.063)	0.556*** (0.282, 0.829)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-2.5 Do Leader Gender, Partisanship, and Disposition Interact?

By varying gender, partisanship, and disposition, some treatment vignettes offer potentially contradictory signals about a leader’s orientation towards conciliation. For instance, a Republican woman president may be perceived as more hawkish than a Democratic woman president. Similarly, a female hawk may be perceived as more conciliatory than a male dove. It is not clear how respondents will interpret conflicting signals from leader gender, leader partisanship, and leader disposition. We study these dynamic in a series of exploratory analyses in Tables A.13 and A.14. Specifically, we calculate gendered peace premia for Democratic and Republican presidents and hawkish and dovish presidents in a regression framework. There are no distinguishable differences in premia across these conditions.

Table A.13: Study 2 — Gendered Peace Premia for Democratic vs. Republican Presidents

	Sample: All Respondents		Sample: Passed Manipulation Check	
	Disapproval (% Points)	Disapproval (7-Point Scale)	Disapproval (% Points)	Disapproval (7-Point Scale)
Democratic Presidents	-2.274 (-11.788, 7.241)	0.002 (-0.389, 0.392)	-5.450 (-19.212, 8.312)	-0.087 (-0.620, 0.447)
Republican Presidents	7.289 (-2.039, 16.618)	0.099 (-0.266, 0.464)	7.714 (-5.295, 20.724)	0.146 (-0.337, 0.629)
Difference in Premia	-9.563 (-22.887, 3.761)	-0.097 (-0.632, 0.437)	-13.164 (-32.102, 5.773)	-0.233 (-0.952, 0.487)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses.

Table A.14: Study 2 — Gendered Peace Premia for Hawkish vs. Dovish Presidents

	Sample: All Respondents		Sample: Passed Manipulation Check	
	Disapproval (% Points)	Disapproval (7-Point Scale)	Disapproval (% Points)	Disapproval (7-Point Scale)
Hawkish Presidents	2.793 (-7.365, 12.951)	0.079 (-0.314, 0.473)	-2.023 (-18.553, 14.507)	-0.021 (-0.601, 0.558)
Dovish Presidents	2.452 (-6.039, 10.943)	0.028 (-0.327, 0.382)	3.519 (-8.529, 15.567)	0.032 (-0.442, 0.505)
Difference in Premia	0.341 (-12.899, 13.580)	0.052 (-0.478, 0.582)	-5.542 (-25.997, 14.913)	-0.053 (-0.801, 0.695)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses.

A-2.6 Mediation Analysis

Consistent with [Mattes and Weeks \(2019\)](#), we find evidence that perceived policy credibility and moderation are significant mediators explaining the dispositional peace premium in Study 2. We also find evidence for a third mechanism discussed in our pre-analysis plan: trustworthiness. Given the strong statistical correlation between policy credibility and trustworthiness ($\rho = 0.69$), it makes sense that trustworthiness would also serve as an important mechanism. Conceptually, if respondents do not believe that President Richards chose the best strategy for dealing with China specifically (policy credibility), then it is also logical for them to not trust President Richards in a general sense (trustworthiness). By contrast, the correlation between perceived moderation and trustworthiness is much smaller ($\rho = 0.13$). Results from causal mediation analyses ([Imai, Keele, and Tingley, 2010](#); [Hicks and Tingley, 2011](#)) support the main findings from Figure 5.

Table A.15: Study 2 — Average Causal Mediation Effects for the Dispositional Peace Premium

	Competence		Moderation		Trustworthiness	
	(1) Hawkish President	(2) Dovish President	(3) Hawkish President	(4) Dovish President	(5) Hawkish President	(6) Dovish President
Average Mediation Effect	0.233 (0.073, 0.395)	0.830 (0.668, 0.992)	0.012 (-0.021, 0.046)	0.065 (0.028, 0.110)	0.141 (0.005, 0.281)	0.2912 (0.171, 0.413)
Average Direct Effect	0.181 (0.061, 0.303)	0.121 (0.011, 0.234)	0.400 (0.205, 0.598)	0.891 (0.715, 1.070)	0.274 (0.129, 0.421)	0.662 (0.526, 0.802)
Total Effect	0.414 (0.375, 0.454)	0.951 (0.896, 1.011)	0.411 (0.251, 0.576)	0.956 (0.812, 1.103)	0.415 (0.398, 0.426)	0.954 (0.921, 0.981)
% of Total Effect Mediated	56.457% (51.338, 62.164)	87.368% (82.072, 92.647)	2.853% (2.039, 4.688)	6.797% (5.900, 8.016)	33.938% (33.118, 35.521)	30.510% (29.685, 31.638)
Respondent-Level Covariates	✓	✓	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓	✓	✓

Note: 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

In Table A.16 we show the key findings of a formal causal mediation analysis for out-partisan presidents. The results suggest that policy credibility is a significant mediator for out-partisan presidents and that this helps explain the gendered peace premium. Policy credibility mediates 55% of the total effect of adopting a conciliatory policy on disapproval for male leaders and 73% for female leaders. These results comport with evidence on the main gendered peace premium identified in Study 1 (Table A.4).

Table A.16: Study 2 — Average Causal Mediation Effect of Policy Credibility for Out-Partisan Presidents

	Out-Partisan Male President	Out-Partisan Female President
Average Mediation Effect	0.326 (0.067, 0.591)	0.648 (0.374, 0.925)
Average Direct Effect	0.265 (0.063, 0.470)	0.245 (0.051, 0.442)
Total Effect	0.590 (0.536, 0.656)	0.892 (0.812, 0.986)
% of Total Effect Mediated	55.330% (49.640, 60.735)	72.767% (65.698, 79.754)
Respondent-Level Covariates	✓	✓
President's Perceived Race	✓	✓

Note: 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-2.7 Heterogeneity in the Gendered Peace Premium

In Table A.17 we test a number of heterogeneous effect hypotheses we had outlined in our pre-analysis plan. We find no evidence of heterogeneous effects, *except* with respect to second-order sexism (column 4). In results omitted for space we find substantively similar evidence with respect to heterogeneity in the out-partisan gendered peace premium. Interestingly, the gendered peace premium is significantly *lower* for those scoring higher in second-order sexism. While respondents high in second-order sexism weakly reward female leaders for pursuing peace, those lower in second-order sexism punish them. In our pre-analysis plan we had expected the opposite to hold. The most logical explanation is that even though respondents high in second-order sexism believe *others* hold sexist views, they do not necessarily hold sexist views themselves—second-order sexism is negatively correlated with hostile sexism. For this reason, those high in second-order sexism may actually want to counter gender stereotypes, rewarding female leaders for pursuing peace. Overall, this finding is ripe for future research.

Table A.17: Study 2 — Heterogeneity in the Gendered Peace Premium

Heterogeneity by Respondent:	Disapproval (7-Point Scale)						
	(1) Partisanship	(2) Hostile Sexism	(3) Benevolent Sexism	(4) Second-Order Sexism	(5) Militant Assertiveness	(6) Education	(7) Gender
Republican Respondent	0.015 (-0.432, 0.461)						
Non-Republican Respondent	0.055 (-0.283, 0.394)						
Difference in Premia	-0.041 (-0.601, 0.520)						
High Sexism Respondent		0.099 (-0.330, 0.528)	-0.124 (-0.657, 0.409)	-0.329 (-0.806, 0.148)			
Low Sexism Respondent		0.069 (-0.490, 0.628)	0.068 (-0.376, 0.513)	0.424 (-0.132, 0.979)			
Difference in Premia		0.030 (-0.677, 0.736)	-0.192 (-0.886, 0.501)	-0.752** (-1.485, -0.021)			
Hawkish Respondent					0.136 (-0.351, 0.623)		
Dovish Respondent					-0.163 (-0.658, 0.332)		
Difference in Premia					0.299 (-0.395, 0.993)		
High Education Respondent						0.151 (-0.251, 0.553)	
Low Education Respondent						-0.361 (-0.896, 0.175)	
Difference in Premia						0.511 (-0.156, 1.179)	
Female Respondent							-0.081 (-0.456, 0.293)
Male Respondent							0.127 (-0.270, 0.523)
Difference in Premia							-0.208 (-0.753, 0.337)
Respondent-Level Covariates	✓	✓	✓	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓	✓	✓	✓
Observations	1970	1324	1309	1189	1321	1415	1970

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

A-2.8 Other Pre-Registered Heterogeneous Effects

In addition to the heterogeneous effect analyses in Table A.17, we also pre-registered a number of additional expectations. First, we anticipated that hostile sexists would be more disapproving of women leaders in general. In column 1 we do not find evidence in support. Second, following from Mattes and Weeks (2019), we expected hawkish respondents to be more disapproving of conciliation irrespective of leader sex. Results in column 2 show hawkish respondents are more likely to disapprove of leaders pursuing conciliation. Because Republicans also tend to be more hawkish (Kertzer and Brutger, 2016), we also expected Republicans to be more disapproving of conciliation irrespective of leader sex. In column 3 we find that this is the case. Finally, column 4 considers whether female respondents are less disapproving of women presidents irrespective of their policies (conciliatory or status quo). We find no evidence for this hypothesis.

Table A.18: Study 2 — Other Pre-Registered Heterogeneous Effects

	Disapproval (7-Point Scale)			
	(1)	(2)	(3)	(4)
Female President × Hostile Sexism	0.014 (-0.138, 0.166)			
Conciliatory Policy × Militant Assertiveness		0.864*** (0.689, 1.040)		
Conciliatory Policy × Republican Respondent			0.166*** (0.106, 0.227)	
Female President × Female Respondent				0.142 (-0.130, 0.415)
Female President	0.042 (-0.431, 0.516)	0.106 (-0.026, 0.239)	0.084 (-0.052, 0.219)	0.011 (-0.186, 0.209)
Conciliatory Policy	0.688*** (0.552, 0.824)	-2.026*** (-2.594, -1.458)	0.044 (-0.222, 0.310)	0.688*** (0.552, 0.824)
Hostile Sexism	-0.055 (-0.170, 0.059)			
Militant Assertiveness		-0.532*** (-0.654, -0.410)		
Republican Respondent			-0.059*** (-0.102, -0.015)	
Female Respondent				0.190* (-0.005, 0.385)
Democratic President	0.148** (0.010, 0.285)	0.139** (0.006, 0.272)	0.154** (0.018, 0.291)	0.148** (0.011, 0.285)
Hawkish President	0.428*** (0.291, 0.565)	0.458*** (0.325, 0.590)	0.433*** (0.297, 0.569)	0.428*** (0.291, 0.565)
Respondent-Level Covariates	✓	✓	✓	✓
President's Perceived Race	✓	✓	✓	✓
Observations	1970	1970	1970	1970

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 95% confidence intervals are in parentheses. Respondent-level covariates are: gender, education, income, age, race, partisanship, militant assertiveness, hostile sexism, benevolent sexism, second-order sexism, and the randomized order in which a respondent received the sexism battery. President's perceived race is each respondent's belief about the race of the president.

In results omitted for space we also consider a suggestion from Mattes and Weeks

(2019, p. 61). Specifically, we explore whether the dispositional peace premium varies based on each respondent’s level of militant assertiveness. Does the advantage hawkish leaders have over dovish leaders in pursuing peace vary based on whether respondents are hawkish or dovish themselves? We find that among hawkish respondents, the dispositional peace premium is modest (effect = 0.382; $p = 0.125$). This premium is large and distinguishable among dovish respondents (effect = 1.512; $p < 0.001$). The difference in these premia (difference = -1.130, $p = 0.002$) suggests dovish respondents actually impose a larger dispositional peace premium than hawkish respondents. At first glance, this effect seems counterintuitive: hawkish respondents should be especially predisposed against conciliation. Indeed, in Table A.18 we find hawks are strongly opposed to conciliation. However, the dispositional peace premium is greater among dovish respondents because they (a) punish hawkish leaders more for maintaining the status quo, and (b) *reward* hawkish (but not dovish) leaders for pursuing conciliation.

A-3 References for Supplementary Materials

1. Aronow, Peter M., Josh Kalla, Lilla Orr, and John Ternovski. 2020. “Evidence of Rising Rates of Inattentiveness on Lucid in 2020.”
2. Bendix, William, and Gyung-Ho Jeong. 2019. “Gender and Foreign Policy: Are Female Members of Congress More Dovish than Their Male Colleagues?” *Political Research Quarterly* 73(1): 126–140.
3. Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. “Information Equivalence in Survey Experiments.” *Political Analysis* 26(4): 399–416.
4. Guisinger, Alexandra and Elizabeth N Saunders. 2017. “Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion across International Issues.” *International Studies Quarterly* 61(2): 425–441.
5. Hicks, Raymond, and Dustin Tingley. 2011. “Causal Mediation Analysis.” *The Stata Journal* 11(4): 605–619.
6. Imai, Kosuke, Luke Keele and Dustin Tingley. 2010. “A General Approach to Causal Mediation Analysis.” *Psychological Methods* 15(4): 309–334.
7. Kertzer, Joshua D, Deborah Jordan Brooks and Stephen G Brooks. 2021. “Do Partisan Types Stop at the Water’s Edge?” *Journal of Politics* 83(4): 1764–1782.
8. Kertzer, Joshua D., and Ryan Brutger. 2016. “Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory.” *American Journal of Political Science* 60(1): 234–249.
9. Mattes, Michaela, and Jessica L.P. Weeks. 2019. “Hawks, Doves, and Peace: An Experimental Approach.” *American Journal of Political Science* 63(1): 53–66.
10. Renshon, Jonathan, Allan Dafoe and Paul Huth. 2018. “Leader Influence and Reputation Formation in World Politics.” *American Journal of Political Science* 62(2): 325–339.

11. Tomz, Michael, Jessica L.P. Weeks, and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74(1):119–143.
12. Trager, Robert F., and Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal of Political Science* 55(3): 526–545.
13. Ziv, Guy. 2014. *Why Hawks Become Doves: Shimon Peres and Foreign Policy Change in Israel*. State University of New York Press.