

Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining*

Joshua A. Schwartz[†] and Christopher W. Blair[‡]

Supplementary Materials

Contents

| | |
|---|-------------|
| Female Leaders in World Politics | A.1 |
| Pre-Registration | A.3 |
| TESS Experiment | A.4 |
| Survey Text | A.4 |
| Robustness | A.8 |
| Heterogeneous Effects | A.13 |
| mTurk Experiment | A.21 |
| Survey Text | A.21 |
| Results | A.26 |

*This is one of several joint articles by the authors; the ordering of names reflects a principle of rotation with equal authorship implied.

[†]PhD Candidate, Department of Political Science, University of Pennsylvania, josha@sas.upenn.edu

[‡]PhD Candidate, Department of Political Science, University of Pennsylvania, cwblair@sas.upenn.edu

Female Leaders in World Politics

Table A.1 lists female heads of state according to three prominent datasets. [Archigos](#) covers the period from 1875-2015, [LEAD](#) covers the period from 1875-2004, and [Jalalzai \(2013\)](#) covers the period from 1960-2010. Post and Sen (2020: 11) note that data from a fourth source, the [Worldwide Guide to Women in Leadership](#), agrees with [Archigos](#).

Table A.1: Female Heads of State

| Leader | Country | Years In Office | Archigos | LEAD | Jalalzai (2013) |
|---|----------------------------|---------------------------------|----------|------|-----------------|
| Empress Dowager Cixi | China | 1861-1908 | | ✓ | |
| Empress Dowager Jonyu | China | 1911-1912 | ✓ | ✓ | |
| Empress Judith Zewditu | Ethiopia | 1916-1930 | | ✓ | |
| Sirimavo Bandaranaike | Sri Lanka | 1960-1965; 1970-1977; 1994-2000 | ✓ | ✓ | ✓ |
| Indira Gandhi | India | 1966-1977; 1980-1984 | ✓ | ✓ | ✓ |
| Golda Meir | Israel | 1969-1974 | ✓ | ✓ | ✓ |
| Isabel Perón | Argentina | 1974-1976 | ✓ | ✓ | ✓ |
| Elisabeth Domitien | Central African Republic | | | | ✓ |
| Maria de Lourdes Pintasilgo | Portugal | 1979-1980 | | | ✓ |
| Margaret Thatcher | United Kingdom | 1979-1990 | ✓ | ✓ | ✓ |
| Lidia Gueiler Tejada | Bolivia | 1979-1980 | ✓ | ✓ | ✓ |
| Eugenia Charles | Dominica | 1980-1995 | | ✓ | ✓ |
| Vigdís Finnbogadóttir | Iceland | 1980-1996 | | | ✓ |
| Gro Harlem Brundtland | Norway | 1981; 1986-1989; 1990-1996 | ✓ | ✓ | ✓ |
| Dzeliwe Shongwe | Swaziland | 1982-1983 | | ✓ | |
| Milka Planinc | Yugoslavia | 1982-1986 | | | ✓ |
| Agatha Barbara | Malta | 1982-1987 | | | ✓ |
| Ntfombi | Swaziland | 1983-1986 | | ✓ | |
| Carmen Pereira | Guinea-Bissau | 1984 | | | ✓ |
| Corazon Aquino | Philippines | 1986-1992 | ✓ | ✓ | ✓ |
| Benazir Bhutto | Pakistan | 1988-1990; 1993-1996 | ✓ | ✓ | ✓ |
| Sabine Bergmann-Pohl | German Democratic Republic | 1990 | | | ✓ |
| Ertha Pascal-Trouillot | Haiti | 1990-1991 | ✓ | ✓ | ✓ |
| Mary Robinson | Ireland | 1990-1997 | | | ✓ |
| Kazimira Prunskienė | Lithuania | 1990-1991 | | | ✓ |
| Violeta Chamorro | Nicaragua | 1990-1997 | ✓ | ✓ | ✓ |
| Khaleda Zia | Bangladesh | 1991-1996; 2001-2006 | ✓ | ✓ | ✓ |
| Édith Cresson | France | 1991-1992 | | | ✓ |
| Hanna Suchocka | Poland | 1992-1993 | | | ✓ |
| Kim Campbell | Canada | 1993 | ✓ | ✓ | ✓ |
| Tansu Çiller | Turkey | 1993-1996 | ✓ | ✓ | ✓ |
| Sylvie Kinigi | Burundi | 1993-1994 | ✓ | ✓ | ✓ |
| Agathe Uwilingiyimana | Rwanda | 1993-1994 | | | ✓ |
| Reneta Indzhova | Bulgaria | 1994-1995 | ✓ | ✓ | ✓ |
| Chandrika Kumaratunga | Sri Lanka | 1994-2005 | ✓ | ✓ | ✓ |
| Claudette Werleigh | Haiti | 1995-1996 | | | ✓ |
| Sheikh Hasina Wazed | Bangladesh | 1996-2001; 2009-Present | ✓ | ✓ | ✓ |
| Ruth Perry | Liberia | 1996-1997 | ✓ | ✓ | |
| Rosalía Arteaga | Ecuador | 1997 | ✓ | ✓ | ✓ |
| Jenny Shipley | New Zealand | 1997-1999 | ✓ | ✓ | ✓ |
| Janet Jagan | Guyana | 1997-1999 | ✓ | ✓ | ✓ |
| Mary McAleese | Ireland | 1997-2011 | | | ✓ |
| Erna Solberg | Norway | 1998 | | ✓ | |
| Ruth Dreifuss | Switzerland | 1999 | ✓ | ✓ | ✓ |
| Nyam-Osoryn Tuyaa | Mongolia | 1999 | | | ✓ |
| Irena Degutienė | Lithuania | 1999 | | | ✓ |
| Mireya Moscoso | Panama | 1999-2004 | ✓ | ✓ | ✓ |
| Helen Clark | New Zealand | | ✓ | ✓ | ✓ |
| Vaira Vīķe-Freiberga | Latvia | 1999-2007 | | | ✓ |
| Tarja Halonen | Finland | 2000-2012 | ✓ | ✓ | |
| Gloria Macapagal Arroyo | Philippines | 2001-2010 | ✓ | ✓ | ✓ |
| Mame Madior Boye | Senegal | 2001-2002 | | | ✓ |
| Megawati Sukarnoputri | Indonesia | 2001-2004 | ✓ | ✓ | ✓ |
| Sang Chang | South Korea | 2002 | | | ✓ |
| Nataša Mičić | Serbia | 2002-2004 | | | ✓ |
| Maria das Neves Ceita Baptista de Sousa | São Tomé and Príncipe | 2002-2004 | | | ✓ |
| Beatriz Merino Lucero | Peru | 2003 | | | ✓ |
| Anneli Jäätteenmäki | Finland | 2003 | | | ✓ |
| Nino Burjanadze | Georgia | 2003-2004 | ✓ | ✓ | ✓ |
| Radmila Šekerinska | Macedonia | 2004 | | ✓ | ✓ |

Table A.1: Female Heads of State, continued

| Leader | Country | In Office | Archigos | LEAD | Jalalzai (2013) |
|--------------------------------|--------------------------|----------------------|----------|------|-----------------|
| Luísa Dias Diogo | Mozambique | 2004-2006 | | | ✓ |
| Cynthia A. Pratt | The Bahamas | 2005 | ✓ | | |
| Yulia Tymoshenko | Ukraine | 2005; 2007-2010 | | | ✓ |
| Ivy Matsepe-Casaburri | South Africa | 2005; 2008 | | | ✓ |
| Maria do Carmo Silveira | São Tomé and Príncipe | 2005-2006 | | | ✓ |
| Angela Merkel | Germany | 2005-Present | ✓ | | ✓ |
| Ellen Johnson Sirleaf | Liberia | 2006-2018 | ✓ | | ✓ |
| Michelle Bachelet | Chile | 2006-2010; 2014-2018 | ✓ | | ✓ |
| Myeong-Sook Han | South Korea | 2006-2007 | | | ✓ |
| Portia Simpson-Miller | Jamaica | 2006-2007; 2012-2016 | ✓ | | ✓ |
| Dalia Itzik | Israel | 2007 | | | ✓ |
| Micheline Calmy-Rey | Switzerland | 2007; 2011 | ✓ | | ✓ |
| Cristina Fernández de Kirchner | Argentina | 2007-2015 | ✓ | | ✓ |
| Pratibha Patil | India | 2007-2012 | | | ✓ |
| Zinaida Greceanii | Moldova | 2008-2009 | ✓ | | ✓ |
| Jóhanna Sigurðardóttir | Iceland | 2009-2013 | ✓ | | ✓ |
| Rose Francine Rogombé | Gabon | 2009 | ✓ | | ✓ |
| Jadranka Kosor | Croatia | 2009-2011 | | | ✓ |
| Dalia Grybauskaitė | Lithuania | 2009-Present | ✓ | | ✓ |
| Doris Leuthard | Switzerland | 2010 | ✓ | | ✓ |
| Mari Kiviniemi | Finland | 2010-2011 | | | ✓ |
| Roza Otunbayeva | Kyrgyzstan | 2010-2011 | ✓ | | |
| Laura Chinchilla Miranda | Costa Rica | 2010-2014 | ✓ | | ✓ |
| Kamla Persad-Bissessar | Trinidad and Tobago | 2010-2015 | ✓ | | ✓ |
| Julia Gillard | Australia | 2010-2013 | ✓ | | ✓ |
| Iveta Radičová | Slovenia | 2010-2012 | ✓ | | ✓ |
| Dilma Rousseff | Brazil | 2011-2016 | ✓ | | |
| Yingluck Shinawatra | Thailand | 2011-2014 | ✓ | | |
| Helle Thorning-Schmidt | Denmark | 2011-2015 | ✓ | | |
| Eveline Widmer-Schlumpf | Switzerland | 2012 | ✓ | | |
| Joyce Hilda Banda | Malawi | 2012-2014 | ✓ | | |
| Park Geun-hye | South Korea | 2013-2017 | ✓ | | |
| Alenka Bratušek | Slovenia | 2013-2014 | ✓ | | |
| Erna Solberg | Norway | 2013-Present | ✓ | | |
| Laimdota Straujuma | Latvia | 2014-2016 | ✓ | | |
| Catherine Samba-Panza | Central African Republic | 2014-2016 | ✓ | | |
| Simonetta Sommaruga | Switzerland | 2015 | ✓ | | |
| Kolinda Grabar-Kitarović | Croatia | 2015-Present | ✓ | | |
| Natalia Gherman | Moldova | 2015 | ✓ | | |

Pre-Registration

Although we conducted an exploratory mTurk pilot, we pre-registered our main study fielded with TESS. The pre-registration was filed with Evidence in Governance and Politics (EGAP). Our pre-registration plan (EGAP #20190731AB) is available [here](#). The following hypotheses were pre-registered:

H_{1A}: Female leaders facing male opponents will pay higher inconsistency costs compared to the MM dyad.

H_{1B}: Female leaders facing female opponents will pay higher inconsistency costs compared to the MM dyad.

H_{2A}: Female leaders facing male opponents will pay lower belligerence costs compared to the MM dyad.

H_{2B}: Female leaders facing female opponents will pay lower belligerence costs compared to the MM dyad.

H_{3A}: Male leaders facing female opponents will pay higher inconsistency costs compared to the MM dyad.

H_{3B}: Male leaders facing female opponents will pay lower belligerence costs compared to the MM dyad.

H₄: The above hypothesized effects should be stronger among older respondents.

H₅: The above hypothesized effects should be stronger among more sexist respondents.

H₆: Democrats will punish leaders more for belligerence, while Republicans will punish leaders more for inconsistency.

H₇: Individuals low in militant assertiveness will punish leaders more for belligerence, while individuals high in militant assertiveness will punish leaders more for inconsistency.

TESS Experiment

Survey Text

Sexism Pre-Test (Question order randomized; red = hostile and black = benevolent; half of respondents received these questions pre-treatment and half post-treatment):

1. Please select the extent to which you agree or disagree with the following statements:

| | Strongly Disagree | Somewhat Disagree | Slightly Disagree | Slightly Agree | Somewhat Agree | Strongly Agree |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Feminists are making reasonable demands of men. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Women must overcome more obstacles than men to be professionally successful. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Women who complain about discrimination often cause more problems than they solve. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| When women demand equality these days, they are actually seeking special favors. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Compared to men, women tend to have a superior moral sensibility. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Many women have a quality of purity that few men possess. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Men have no special obligation to provide financially for the women in their lives. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| There is no need for men to cherish or protect women. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Foreign Policy Dispositions Pre-Test (Question order randomized):

2. Please select the extent to which you agree or disagree with the following statements:

| | Strongly Disagree | Somewhat Disagree | Neither Agree Nor Disagree | Somewhat Agree | Strongly Agree |
|--|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|
| The best way to ensure world peace is through American military strength. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Going to war is unfortunate, but sometimes the only solution to international problems. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| The use of military force only makes problems worse. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Random Assignment to Scenario:

The following questions are about US relations with other countries around the world. You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach US leaders could take in the future and ask whether you approve or disapprove.

Treatment Screen: [1 of 24 scenarios—equal assignment probability; names randomized; blocked on respondent party identification]

Stay Out Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] U.S. President, [Erica/Eric, Stephanie/Steven] Smith, says the United States will stay out of the conflict. The attacking country continues to invade. In the end, [Erica/Eric, Stephanie/Steven] Smith decides not to send troops, and the attacking country gains 20% of the contested territory.

To summarize:

- The attacking country is led by a [female/male] leader.
- The US President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith.
- [Erica/Eric, Stephanie/Steven] Smith says the United States will stay out of the conflict.
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does not send troops.
- The attacking country gains 20% of the contested territory.

Not Engage Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] U.S. President, [Erica/Eric, Stephanie/Steven] Smith, says that if the attack continues, the US military will push out the invaders. The attacking country continues to invade. In the end, [Erica/Eric, Stephanie/Steven] Smith does not send troops, and the attacking country gains 20% of the contested territory.

To summarize:

- The attacking country is led by a [female/male] leader.
- The US President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith.
- [Erica/Eric, Stephanie/Steven] Smith says that if the attack continues, the United States military will push out the invaders.
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does not send troops.
- The attacking country gains 20% of the contested territory.

Engage Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] U.S. President, [Erica/Eric, Stephanie/Steven] Smith, says that if the attack continues, the US military will push out the invaders. The attacking country continues to invade. In the end, [Erica/Eric, Stephanie/Steven] Smith orders the military to engage. The attacking country gains 20% of the contested territory, and the US experiences zero casualties.

To summarize:

- The attacking country is led by a [female/male] leader.
- The US President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith.
- [Erica/Eric, Stephanie/Steven] Smith says that if the attack continues, the United States military will push out the invaders.
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does decide to send troops.
- The attacking country ultimately gains 20% of the contested territory.
- The US experiences zero casualties.

Dependent Variable:

3. To what extent do you disapprove or approve of how the US President handled the situation?

| | | | | | | |
|------------------------|-----------------------|-----------------------------|-----------------------------------|--------------------------|-----------------------|-----------------------|
| Strongly Disapprove | Disapprove | Lean Toward Disapproving | Neither Approve Nor Disapprove | Lean Toward Approving | Approve | Strongly Approve |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

4. Please write down four words that you believe describe the US President in this situation.

| | | | |
|--|--|--|--|
| | | | |
|--|--|--|--|

Placebo & Manipulation Check Questions:

5. What is your best estimate of how democratic the attacking country is, on a scale of -10 to +10, where -10 is the least democratic and +10 is the most democratic? Next to the options are some example countries to help calibrate your answer.:

| | | | | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| -10 (North Korea) | -8 | -6 (China) | -4 | -2 (Sudan) | 0 | 2 (Tanzania) | 4 | 6 (Colombia) | 8 | 10 (Canada) |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

6. What is your best guess of what race the US President is?

| | | | | |
|-----------------------|----------------------------|-----------------------|-----------------------|-----------------------|
| Caucasian/ White | African American/ Black | Asian | Hispanic/ Latino | Multi-Racial |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

7. What was the name of the US President?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Eric | Erica | Steven | Stephanie | None of the Above |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

8. What was the sex of the leader of the attacking country?

| | | |
|-----------------------|-----------------------|-----------------------|
| Male | Female | Not Stated |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

9. What was the party affiliation of the U.S. President?

Republican Democrat Not Stated



Robustness

Table A.2 replicates Table 3 in the main text; however, whereas Table 3 in the main text excludes respondents who failed the attention check, Table A.2 includes all respondents.

Table A.2: Percentage Point Difference in Mean Disapproval Compared to the Male-Male Baseline (All Respondents)

| Dyad (Baseline = Male-Male) | Audience Cost (%) (1) | Inconsistency Cost (%) (2) | Belligerence Cost (%) (3) |
|--------------------------------|-----------------------------|----------------------------------|---------------------------------|
| Female-Male | +8.2 (27.3 – 19.1) | +19.2*** (52.0 – 32.9) | -11.0* (-24.7 – -13.7) |
| Female-Female | +5.3 (24.4 – 19.1) | +17.1*** (50.0 – 32.9) | -11.9** (-25.6 – -13.7) |
| Male-Female | -0.6 (18.6 – 19.1) | +10.8* (43.6 – 32.9) | -11.3* (-25.1 – -13.7) |

Note: Results depict average treatment effects (ATE) for a binary measure of disapproval calculated from 2,000 bootstraps. The main quantities reflect the average percentage point difference in disapproval for the dyad compared to the male-male baseline. For example, 19.2 percentage points more of respondents disapprove of a female president acting inconsistently against a foreign male leader than a male president acting inconsistently against a foreign male leader. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$. The mean disapproval for the two experimental groups used to calculate ATE are in parentheses.

Table A.3 utilizes the full 7-point measure of approval or disapproval instead of the binary measure employed in Table 3 of the main text. Table A.3 only examines respondents that passed the attention check. All results are robust to using the 7-point rather than binary measure, though the difference in belligerence costs between the MF and MM dyad is estimated less precisely ($p \approx 0.128$).

Table A.3: Difference in Mean Disapproval on a 7-Point Scale Compared to the Male-Male Baseline (Excluding Respondents that Failed the Attention Check)

| Dyad (Baseline = Male-Male) | Audience Cost (1) | Inconsistency Cost (2) | Belligerence Cost (3) |
|--------------------------------|-------------------------|------------------------------|-----------------------------|
| Female-Male | +0.09 (0.99 – 0.90) | +0.56*** (2.17 – 1.61) | -0.46** (-1.17 – -0.71) |
| Female-Female | -0.01 (0.89 – 0.90) | +0.35* (1.95 – 1.61) | -0.36* (-1.06 – -0.71) |
| Male-Female | +0.28 (1.18 – 0.90) | +0.56** (2.16 – 1.61) | -0.28 (-0.99 – -0.71) |

Note: Results depict average treatment effects (ATE) for a 7-point measure of disapproval calculated from 2,000 bootstraps. The main quantities reflect the average difference in disapproval on a 7-point scale for the dyad compared to the male-male baseline. For example, average disapproval is 0.56 points higher on a 7-point scale for a female president that acts inconsistently against a foreign male leader compared to a male president that acts inconsistently against a foreign male leader. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$. The mean disapproval for the two experimental groups used to calculate ATE are in parentheses.

Table A.4 also utilizes the full 7-point measure of approval or disapproval instead of the binary measure employed in Table 3, and includes all respondents. All of the results are robust to this change, though the difference in inconsistency costs between the FF and MM dyads is estimated less precisely ($p \approx 0.197$), and the difference in belligerence costs between the MF and MM dyads is just over conventional levels of significance ($p \approx 0.104$).

Table A.4: Difference in Mean Disapproval on a 7-Point Scale Compared to the Male-Male Baseline (All Respondents)

| Dyad (Baseline = Male-Male) | Audience Cost (1) | Inconsistency Cost (2) | Belligerence Cost (3) |
|--------------------------------|-------------------------|------------------------------|-----------------------------|
| Female-Male | -0.04 (0.76 – 0.80) | +0.35* (1.72 – 1.37) | -0.39** (-0.96 – -0.57) |
| Female-Female | -0.13 (0.67 – 0.80) | +0.19 (1.56 – 1.37) | -0.33* (-0.90 – -0.57) |
| Male-Female | +0.06 (0.86 – 0.80) | +0.35* (1.72 – 1.37) | -0.29 (-0.86 – -0.57) |

Note: Results depict average treatment effects (ATE) for a 7-point measure of disapproval calculated from 2,000 bootstraps. The main quantities reflect the average difference in disapproval on a 7-point scale for the dyad compared to the male-male baseline. For example, average disapproval is 0.35 points higher on a 7-point scale for a female president that acts inconsistently against a foreign male leader compared to a male president that acts inconsistently against a foreign male leader. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$. The mean disapproval for the two experimental groups used to calculate ATE are in parentheses.

Table A.5 tests our hypotheses in a regression context. Models 1 and 3 report results excluding respondents that failed the attention check, and models 2 and 4 report results with all respondents. Recall that audience costs equal disapproval in the not engage condition minus disapproval in the stay out condition (this is equivalent to inconsistency plus belligerence costs); inconsistency costs equal disapproval in the not engage condition minus disapproval in the engage condition; and belligerence costs equal disapproval in the engage condition minus disapproval in the stay out condition. In models 1 and 2, the inconsistency costs for the MM dyad equal the coefficient on MM Not Engage. The following calculation generates inconsistency costs for the other dyads:

$$\text{Inconsistency Costs} = [\text{Gender Category}] \text{ Not Engage} - [\text{Gender Category}] \text{ Engage}$$

In models 3 and 4, the audience costs for the MM dyad equal the coefficient on MM Not Engage and the belligerence costs for the MM dyad equal the coefficient on MM Engage. The following calculation can be used to generate the audience and belligerence costs for the other dyads:

$$\text{Audience Costs} = [\text{Gender Category}] \text{ Not Engage} - [\text{Gender Category}] \text{ Stay Out}$$

$$\text{Belligerence Costs} = [\text{Gender Category}] \text{ Engage} - [\text{Gender Category}] \text{ Stay Out}$$

To determine whether there are statistically significant differences between the three dyads and the MM control, we derive parameter estimates from 2,000 bootstrapped runs of each of our regression models. This is the same strategy employed by Kertzer and Brutger (2016). The quantities of interest are thus not the regression coefficients, but the differences between the MM dyad and the FM, MF, and FF dyads reported in the bottom rows of the table. P-values indicate whether the results are statistically greater than zero for inconsistency costs, or less than zero for belligerence costs. Note that the regression results are substantively similar to results in Table 3.

Table A.5: Regression Results

| | Inconsistency | | Belligerence | |
|------------------------|-------------------|--------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| MM Stay Out | 0.70*** (0.16) | 0.59*** (0.15) | | |
| MM Not Engage | 1.60*** (0.17) | 1.35*** (0.15) | 0.90*** (0.16) | 0.76*** (0.15) |
| MM Engage | | | -0.70*** (0.16) | -0.59*** (0.15) |
| FM Stay Out | 1.02*** (0.17) | 0.82*** (0.16) | 0.32** (0.16) | 0.23 (0.16) |
| FM Not Engage | 1.94*** (0.16) | 1.52*** (0.15) | 1.25*** (0.16) | 0.93*** (0.15) |
| FM Engage | -0.25 (0.16) | -0.19 (0.15) | -0.94*** (0.16) | -0.78*** (0.15) |
| MF Stay Out | 0.87*** (0.16) | 0.82*** (0.15) | 0.17 (0.16) | 0.23 (0.15) |
| MF Not Engage | 2.07*** (0.16) | 1.73*** (0.15) | 1.37*** (0.16) | 1.15*** (0.15) |
| MF Engage | -0.13 (0.17) | -0.07 (0.15) | -0.83*** (0.16) | -0.66*** (0.15) |
| FF Stay Out | 0.94*** (0.16) | 0.85*** (0.15) | 0.24 (0.16) | 0.27* (0.15) |
| FF Not Engage | 1.82*** (0.16) | 1.54*** (0.15) | 1.12*** (0.16) | 0.95*** (0.15) |
| FF Engage | -0.11 (0.17) | -0.03 (0.16) | -0.80*** (0.17) | -0.62*** (0.15) |
| Party US President | 0.07 (0.07) | 0.08 (0.06) | 0.07 (0.07) | 0.08 (0.06) |
| Hostile Sexism | 0.29* (0.17) | 0.44*** (0.16) | 0.29* (0.17) | 0.44*** (0.16) |
| Benevolent Sexism | 0.53*** (0.19) | 0.72*** (0.18) | 0.53*** (0.19) | 0.72*** (0.18) |
| Militant Assertiveness | 0.50*** (0.16) | 0.25 (0.15) | 0.50*** (0.16) | 0.25 (0.15) |
| Party ID | 0.01 (0.02) | 0.06*** (0.02) | 0.01 (0.02) | 0.06*** (0.02) |
| Gender | -0.03 (0.07) | 0.02 (0.07) | -0.03 (0.07) | 0.02 (0.07) |
| Age | 0.07** (0.03) | 0.03 (0.03) | 0.07** (0.03) | 0.03 (0.03) |
| Education | 0.003 (0.04) | -0.002 (0.03) | 0.003 (0.04) | -0.002 (0.03) |
| Income | -0.05* (0.03) | -0.06** (0.03) | -0.05* (0.03) | -0.06** (0.03) |
| Sexism Order | -0.05 (0.07) | -0.05 (0.06) | -0.05 (0.07) | -0.05 (0.06) |
| Regime Placebo | -0.01** (0.01) | -0.02*** (0.01) | -0.01** (0.01) | -0.02*** (0.01) |
| Race Placebo | -0.09 (0.09) | 0.04 (0.08) | -0.09 (0.09) | 0.04 (0.08) |
| Constant | 2.82*** (0.29) | 2.79*** (0.27) | 3.52*** (0.29) | 3.37*** (0.27) |
| FM vs. MM Dyad: | 0.59*** | 0.36* | -0.56** | -0.42** |
| MF vs. MM Dyad: | 0.59** | 0.46** | -0.30 | -0.31* |
| FF vs. MM Dyad: | 0.32* | 0.22 | -0.34* | -0.30* |
| Observations | 1,777 | 2,275 | 1,777 | 2,275 |

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

Heterogeneous Effects

Do the effects of gender stereotypes on audience costs vary across respondent sub-groups? Our analysis of heterogeneous effects focuses on five respondent characteristics—militant assertiveness, partisanship, sexism, age, and respondent gender—that existing literature suggests are relevant in the context of disaggregated audience costs (militant assertiveness and partisanship per Kertzer and Brutger 2016) or that associate with gender stereotypical attitudes (sexism, age, and gender per Ellemers 2018). We pre-registered hypotheses pertinent to these sub-samples (see the pre-registration section for more details on our pre-analysis plan). Specifically, we predicted that irrespective of leader gender, Democrats and those low in militant assertiveness would impose higher belligerence costs, while Republicans and those high in militant assertiveness would impose higher inconsistency costs. These findings are well-established benchmarks in analyses of disaggregated audience costs, so replicating them builds confidence in our design (Kertzer and Brutger 2016). Likewise, because sexism and age in particular are important correlates of gender stereotypical attitudes, we expected the strongest results among more sexist and older respondents—the sub-samples most likely to hold the traditional gender stereotypes underlying the Lack of Fit model. Finally, we also conduct exploratory analyses to more thoroughly probe the intersection of partisanship, respondent gender, and gender stereotypes. Existing literature offers mixed guidance about whether the effects of partisanship supersede gender stereotypes or whether gender has an independent effect (Sanbonmatsu and Dolan 2009; Hayes 2011; Bauer 2017). So, our exploratory analyses looking at the intersection of gender and party speak to this debate.

Table A.6 tests whether Kertzer and Brutger’s (2016) finding that individuals that are high in militant assertiveness and are stronger Republicans will punish leaders more for inconsistency and subjects that are low in militant assertiveness and are stronger Democrats will punish leaders more for belligerence holds in the context of our study. Models 1 and 3 exclude respondents that failed the attention check and models 2 and 4 include all subjects. Following Kertzer and Brutger (2016), we define low and high levels of each dispositional characteristic using the interquartile range, thus comparing individuals in the bottom and top 25%.

The quantities of interest are again not the regression coefficients, but the inconsistency fractions in the bottom rows of the table derived from 2,000 bootstrapped versions of each of our regression models. This measure reflects what fraction of audience costs comes from inconsistency costs, and p-values indicate whether the difference between low and high levels of the dispositional characteristic are significantly different in the expected direction. Per Kertzer and Brutger’s (2016) argument, the inconsistency fraction should be greater for subjects that score higher in militant assertiveness and are stronger Republicans. Results confirm this expectation.

Table A.6: Replication of Kertzer and Brutger 2016

| | Militant Assertiveness | | Party ID | |
|-------------------------|---|---|--|--|
| | (1) | (2) | (3) | (4) |
| Not Engage x Mil Assert | -0.44* | -0.25 | | |
| | (0.24) | (0.23) | | |
| Engage x Mil Assert | -2.36*** | -2.04*** | | |
| | (0.25) | (0.24) | | |
| Not Engage x Republican | | | -0.08 | 0.01 |
| | | | (0.17) | (0.16) |
| Engage x Republican | | | -0.92*** | -0.82*** |
| | | | (0.17) | (0.16) |
| Not Engage | 1.22*** | 0.94*** | 1.02*** | 0.77*** |
| | (0.20) | (0.19) | (0.10) | (0.09) |
| Engage | 0.40** | 0.31 | -0.63*** | -0.52*** |
| | (0.20) | (0.19) | (0.10) | (0.10) |
| Militant Assertiveness | 1.11*** | 0.77*** | | |
| | (0.18) | (0.18) | | |
| Republican | | | 0.40*** | 0.14 |
| | | | (0.13) | (0.12) |
| Gender Opponent | 0.15 | 0.18** | 0.08 | 0.13** |
| | (0.10) | (0.09) | (0.07) | (0.06) |
| Gender US President | 0.11 | 0.05 | 0.04 | 0.01 |
| | (0.09) | (0.09) | (0.07) | (0.06) |
| Party US President | 0.26*** | 0.21** | 0.07 | 0.08 |
| | (0.09) | (0.09) | (0.07) | (0.06) |
| Hostile Sexism | 0.35 | 0.44** | 0.21 | 0.35** |
| | (0.23) | (0.22) | (0.17) | (0.16) |
| Benevolent Sexism | 0.53** | 0.67*** | 0.54*** | 0.72*** |
| | (0.26) | (0.25) | (0.19) | (0.18) |
| Party ID | 0.02 | 0.05* | | |
| | (0.03) | (0.03) | | |
| Militant Assertiveness | | | 0.35** | 0.11 |
| | | | (0.16) | (0.15) |
| Gender | -0.12 | -0.02 | -0.04 | 0.02 |
| | (0.10) | (0.09) | (0.07) | (0.06) |
| Age | 0.11** | 0.10** | 0.07** | 0.03 |
| | (0.05) | (0.04) | (0.03) | (0.03) |
| Education | 0.01 | -0.001 | 0.01 | 0.01 |
| | (0.05) | (0.05) | (0.04) | (0.03) |
| Income | -0.01 | -0.01 | -0.06** | -0.06** |
| | (0.04) | (0.04) | (0.03) | (0.03) |
| Sexism Order | -0.07 | -0.04 | -0.06 | -0.06 |
| | (0.10) | (0.09) | (0.07) | (0.06) |
| Regime Placebo | -0.02** | -0.03*** | -0.01** | -0.02*** |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Race Placebo | -0.13 | -0.03 | -0.09 | 0.04 |
| | (0.12) | (0.11) | (0.09) | (0.08) |
| Constant | 2.82*** | 2.82*** | 3.66*** | 3.80*** |
| | (0.39) | (0.37) | (0.23) | (0.21) |
| Inconsistency Fraction | Low Mil Assert: 0.68*** High Mil Assert: 3.59*** | Low Mil Assert: 0.68*** High Mil Assert: 3.66*** | Democrats: 1.63*** Republicans: 2.70*** | Democrats: 1.70*** Republicans: 2.80*** |
| Observations | 936 | 1,143 | 1,787 | 2,291 |

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

Tables A.7 and A.8 test whether our hypotheses are moderated by sexism, age, partisanship, or respondent gender. Extant research suggests that sexism and age are core factors predicting gender stereotypical attitudes, so we anticipated stronger effects among these sub-groups. To measure sexism, we use Winter’s (2018) eight-question index of benevolent and hostile sexism based on the sexism inventory from Glick and Fiske (1996), and estimate separate models for benevolent and hostile sexists. We standardize each sexism index and split the sample into benevolent and hostile sexists along the interquartile range, where sexists are defined as respondents scoring in the top quartile of the respective index. Specifically, those in the bottom quartile are assigned a value of zero, those in the top quartile receive a value of one, and the middle two quartiles are treated as missing. We also define older individuals as those in the top quartile of age, and younger individuals as those in the bottom quartile. Odd-numbered models exclude subjects that failed the attention check and even-numbered models include all respondents.

To calculate the quantities of interest, we take the difference in inconsistency/belligerence costs for the FM, MF, and FF dyads compared to the MM dyad for respondents in the top quartile of the potential moderator and those in the bottom quartile. For example, when examining the effect of age on inconsistency costs for the FM dyad, we do the following calculation: (FM Inconsistency Costs Old — MM Inconsistency Costs Old) — (FM Inconsistency Costs Young — MM Inconsistency Costs Young). If older respondents punish female leaders more for inconsistency, then this quantity should be greater than zero. Utilizing the results from Table A.7, the difference in inconsistency costs compared to the MM baseline for the relevant dyad can be calculated as follows:

$$\text{Bottom Quartile} = ([\text{Gender Category}] \text{ Not Engage} - [\text{Gender Category}] \text{ Engage}) - (\text{MM Not Engage})$$

$$\text{Top Quartile} = ([\text{Gender Category}] \text{ Not Engage} - [\text{Gender Category}] \text{ Engage}) + ([\text{Gender Category}] \text{ Not Engage} * \text{Moderator} - [\text{Gender Category}] \text{ Engage} * \text{Moderator}) - (\text{MM Not Engage} + \text{MM Not Engage} * \text{Moderator})$$

The difference between these two values is then our quantity of interest, which reduces down to the following:

$$\text{Difference between Top and Bottom Quartiles} = ([\text{Gender Category}] \text{ Not Engage} * \text{Moderator} - [\text{Gender Category}] \text{ Engage} * \text{Moderator}) - (\text{MM Not Engage} * \text{Moderator})$$

Employing the same method for belligerence costs, the difference in belligerence costs compared to the MM baseline for the relevant dyad can be calculated as follows using results from Table A.8:

$$\text{Bottom Quartile} = ([\text{Gender Category}] \text{ Engage} - [\text{Gender Category}] \text{ Stay Out}) - (\text{MM Engage})$$

$$\text{Top Quartile} = ([\text{Gender Category}] \text{ Engage} - [\text{Gender Category}] \text{ Stay Out}) + ([\text{Gender Category}] \text{ Engage} * \text{Moderator} - [\text{Gender Category}] \text{ Stay Out} * \text{Moderator}) - (\text{MM Engage} + \text{MM Engage} * \text{Moderator})$$

The difference between these two values is then our quantity of interest, which reduces down to the following:

*Difference between Top and Bottom Quartiles = ([Gender Category] Engage*Moderator — [Gender Category] Stay Out*Moderator) — (MM Engage*Moderator)*

P-values in Tables [A.7](#) and [A.8](#) indicate whether the difference between the top and bottom quartiles are significantly different in the expected direction.

Tables [A.7](#) and [A.8](#) show that, contrary to our expectations, no significant differences emerge when we evaluate distinct respondent sub-samples.

Table A.7: Heterogeneous Effects — Inconsistency Costs

| Sub-Sample | Benevolent Sexism | | | Hostile Sexism | | | Party ID | | | Respondent Gender | | |
|----------------------------|-------------------|---------|---------|----------------|----------|---------|----------|----------|---------|-------------------|--|--|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | | |
| MM Stay Out x Sub-Sample | 0.37 | 0.31 | 0.81** | 0.41 | 0.62 | 0.38 | -1.44** | -1.41*** | -0.26 | -0.29 | | |
| MM Not Engage x Sub-Sample | -0.02 | -0.14 | 0.40 | 0.12 | 0.61 | 0.64* | -0.88** | -0.89** | (0.32) | (0.30) | | |
| FM Stay Out x Sub-Sample | 0.31 | 0.44 | 1.97*** | 0.74* | 0.67 | 0.37 | -0.37 | -0.34** | -0.07** | -0.39* | | |
| FM Not Engage x Sub-Sample | -0.11 | 0.04 | 0.45 | 0.42 | 0.42 | 0.39 | -0.44 | -0.41** | (0.31) | (0.31) | | |
| FM Engage x Sub-Sample | -0.05 | -0.09 | 0.44 | 0.54 | 0.58 | 0.70 | -1.75*** | -1.60*** | -0.50 | -0.43 | | |
| MF Stay Out x Sub-Sample | 0.60 | 0.43 | 0.42 | 0.25 | 0.75* | 0.37 | -0.12 | -0.31 | -0.43 | -0.60* | | |
| MF Not Engage x Sub-Sample | -0.20 | -0.11 | 0.43 | 0.28 | 1.08*** | 0.69* | -1.47*** | -1.18*** | 0.19 | 0.11 | | |
| MF Engage x Sub-Sample | 0.20 | 0.13 | 0.42 | 0.39 | 0.91** | 0.92** | -0.60 | -0.65 | -0.59* | -0.53* | | |
| FF Stay Out x Sub-Sample | 0.35 | 0.37 | 0.44 | 0.41 | 0.99** | 0.50 | -0.85** | -0.89 | -0.33 | -0.01 | | |
| FF Not Engage x Sub-Sample | -0.33 | -0.11 | 0.42 | 0.42 | 1.09*** | 0.37 | -1.38*** | -1.08*** | -0.27 | -0.27 | | |
| FF Engage x Sub-Sample | 0.11 | 0.21 | 0.41 | 0.28 | 0.40 | 0.43 | -1.66*** | -1.27*** | -0.53* | -0.30 | | |
| MM Stay Out | 0.37 | 0.37 | 0.43 | 0.15 | 0.22 | 0.20 | 1.59*** | 1.48*** | 0.82*** | 0.72*** | | |
| MM Not Engage | 1.54*** | 1.47*** | 1.28*** | 1.14*** | 1.10*** | 0.83** | 2.19*** | 1.97*** | 1.94*** | 1.68*** | | |
| FM Stay Out | 0.40 | 0.40 | 0.22 | 0.32 | 0.45 | 0.32 | 1.50** | 1.79*** | 1.16** | 0.91*** | | |
| FM Not Engage | 1.80*** | 1.40*** | 1.28*** | 1.18*** | 1.39*** | 0.79*** | 2.89*** | 2.39*** | 2.19*** | 1.74*** | | |
| FM Engage | 0.38 | 0.38 | 0.31 | 0.49 | 0.44 | 0.29 | 0.32 | 0.32 | (0.31) | (0.22) | | |
| MF Stay Out | 0.54* | 0.50** | 0.22 | 0.38 | 0.34 | 0.31 | 1.71*** | 1.64*** | 0.77*** | 0.77*** | | |
| MF Not Engage | 2.11*** | 1.74*** | 1.70*** | 1.63*** | 1.41*** | 1.09*** | 2.47*** | 2.26*** | 2.35*** | 2.00*** | | |
| MF Engage | -0.31 | -0.14 | -0.43 | -0.28 | -0.89*** | -0.68** | 0.46 | 0.21 | -0.05 | -0.05 | | |
| FF Stay Out | 0.66** | 0.68** | 0.47 | 0.68** | 0.31 | 0.27 | 1.74*** | 1.49*** | 1.16*** | 0.89*** | | |
| FF Not Engage | 1.85*** | 1.54*** | 0.89*** | 1.01*** | 1.35*** | 1.14*** | 2.68*** | 2.28*** | 2.08*** | 1.73*** | | |
| FF Engage | -0.16 | -0.06 | 0.31 | 0.30 | 0.29 | 0.27 | 0.08 | 0.22 | -0.01 | -0.03 | | |
| Sub-Sample | 0.03 | 0.11 | -0.59* | -0.42 | -0.58** | -0.43 | 1.02*** | 1.17** | 0.30 | 0.29 | | |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Constant | 3.43*** | 3.23*** | 3.67*** | 3.32*** | 3.62*** | 3.53*** | 2.50*** | 2.42*** | 2.65*** | 2.66*** | | |
| Observations | 972 | 1,201 | 1,102 | 1,393 | 1,195 | 1,544 | 1,107 | 1,382 | 1,777 | 2,275 | | |

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

Table A.8: Heterogeneous Effects — Belligerence Costs

| Sub-Sample | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | | |
|----------------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|--------------------|--------------------|--------------------|--------------------|--|--|
| | Benevolent Section | | | Hostile Section | | | Party ID | | | Resistant Gender | | |
| | Age | | | Age | | | Age | | | Age | | |
| MM Not Engage x Sub-Sample | -0.38 (0.43) | -0.45 (0.43) | -0.41 (0.41) | -0.29 (0.39) | -0.01 (0.39) | 0.26 (0.36) | 0.47 (0.40) | 0.52 (0.39) | -0.41 (0.36) | -0.30 (0.30) | | |
| MM Engage x Sub-Sample | -0.37 (0.44) | -0.31 (0.42) | -0.81** (0.41) | -0.41 (0.38) | -0.62 (0.40) | -0.38 (0.36) | 1.45*** (0.41) | 1.41*** (0.36) | 0.26 (0.32) | 0.29 (0.30) | | |
| FM Stay Out x Sub-Sample | -0.03 (0.45) | 0.23 (0.43) | 0.58 (0.42) | 0.33 (0.40) | -0.05 (0.41) | 0.01 (0.38) | 0.14 (0.42) | 0.08 (0.41) | -0.05 (0.31) | 0.11 (0.31) | | |
| FM Not Engage x Sub-Sample | -0.48 (0.47) | -0.27 (0.46) | 0.37 (0.41) | 0.13 (0.39) | -0.04 (0.40) | 0.32 (0.37) | -0.30 (0.41) | -0.19 (0.39) | -0.24 (0.32) | -0.11 (0.31) | | |
| FM Engage x Sub-Sample | -0.41 (0.45) | -0.40 (0.44) | 0.41 (0.38) | 0.39 (0.38) | 0.02 (0.39) | 0.02 (0.35) | 1.11*** (0.41) | 1.11*** (0.37) | 0.02 (0.32) | 0.01 (0.30) | | |
| MF Stay Out x Sub-Sample | 0.23 (0.45) | 0.12 (0.40) | -0.04 (0.38) | -0.13 (0.38) | 0.47 (0.42) | 0.30 (0.37) | -0.01 (0.41) | 0.24 (0.39) | 0.45 (0.32) | 0.40 (0.30) | | |
| MF Not Engage x Sub-Sample | -0.56 (0.47) | -0.42 (0.43) | -0.38 (0.40) | -0.44 (0.38) | 0.30 (0.40) | 0.54 (0.36) | 0.85** (0.42) | 0.70* (0.40) | -0.34 (0.32) | -0.24 (0.30) | | |
| MF Engage x Sub-Sample | -0.17 (0.47) | -0.18 (0.44) | -0.38 (0.41) | -0.22 (0.39) | 0.38 (0.36) | 0.12 (0.33) | 0.60 (0.41) | 1.02*** (0.41) | 0.13 (0.33) | 0.28 (0.30) | | |
| FF Stay Out x Sub-Sample | 0.16 (0.45) | 0.06 (0.42) | -0.33 (0.40) | -0.58 (0.38) | 0.44 (0.39) | -0.01 (0.35) | 0.10 (0.40) | 0.33 (0.39) | -0.20 (0.32) | 0.01 (0.30) | | |
| FF Not Engage x Sub-Sample | 0.01 (0.43) | 0.12 (0.41) | 0.14 (0.39) | 0.14 (0.37) | 0.21 (0.38) | 0.21 (0.35) | 0.21 (0.41) | 0.27 (0.38) | 0.21 (0.30) | 0.20 (0.30) | | |
| FF Engage x Sub-Sample | -0.29 (0.48) | -0.10 (0.44) | -0.46 (0.41) | -0.13 (0.39) | -0.19 (0.41) | -0.31 (0.38) | 1.19*** (0.42) | 1.08*** (0.40) | 0.02 (0.33) | 0.23 (0.31) | | |
| MM Not Engage | 1.18*** (0.34) | 1.10*** (0.33) | 1.25*** (0.28) | 1.00*** (0.27) | 0.88*** (0.30) | 0.62** (0.27) | 0.59** (0.30) | 0.40 (0.29) | 1.13*** (0.24) | 0.93*** (0.23) | | |
| MM Engage | -0.37 (0.29) | -0.37 (0.29) | -0.02 (0.29) | -0.15 (0.28) | -0.22 (0.28) | -0.20 (0.27) | -1.59*** (0.30) | -1.48*** (0.29) | -0.82*** (0.23) | -0.73*** (0.22) | | |
| FM Stay Out | 0.22 (0.31) | 0.03 (0.30) | 0.19 (0.27) | 0.17 (0.27) | 0.23 (0.32) | 0.11 (0.29) | 0.31 (0.32) | 0.31 (0.31) | 0.35 (0.22) | 0.18 (0.22) | | |
| FM Not Engage | 1.44*** (0.30) | 1.03*** (0.29) | 1.25*** (0.28) | 1.03*** (0.28) | 1.17*** (0.30) | 0.77*** (0.29) | 1.29*** (0.30) | 1.02*** (0.29) | 1.37*** (0.22) | 1.09*** (0.22) | | |
| FM Engage | 0.69** (0.32) | 0.57* (0.30) | 0.71** (0.28) | 0.54** (0.28) | 0.65** (0.34) | 0.45** (0.29) | 1.00*** (0.30) | 0.77** (0.29) | 0.82*** (0.23) | 0.62*** (0.22) | | |
| MF Stay Out | 0.18 (0.30) | 0.22 (0.28) | 0.19 (0.29) | 0.23 (0.28) | -0.06 (0.31) | 0.11 (0.29) | 0.11 (0.30) | 0.16 (0.29) | -0.04 (0.23) | 0.04 (0.21) | | |
| MF Not Engage | 1.77*** (0.31) | 1.37*** (0.29) | 1.68*** (0.28) | 1.48*** (0.27) | 1.19*** (0.31) | 0.79*** (0.27) | 0.87*** (0.30) | 0.78*** (0.29) | 1.54*** (0.23) | 1.27*** (0.22) | | |
| MF Engage | -0.68** (0.34) | -0.52* (0.31) | -0.46 (0.29) | -0.43 (0.29) | -1.10*** (0.30) | -0.88** (0.27) | -1.13*** (0.30) | -1.27*** (0.29) | -0.87*** (0.22) | -0.78*** (0.21) | | |
| FF Stay Out | 0.29 (0.31) | 0.31 (0.28) | 0.29 (0.29) | 0.48* (0.28) | -0.07 (0.30) | 0.27 (0.26) | 0.14 (0.30) | 0.01 (0.29) | 0.35 (0.23) | 0.26 (0.21) | | |
| FF Not Engage | 1.47*** (0.30) | 1.19*** (0.28) | 1.49*** (0.29) | 1.06*** (0.28) | 1.14*** (0.30) | 0.66** (0.27) | 1.06*** (0.30) | 1.02*** (0.29) | 1.28*** (0.22) | 1.06*** (0.21) | | |
| FF Engage | -0.53 (0.32) | -0.43 (0.28) | -0.30 (0.29) | -0.42 (0.28) | -0.64** (0.32) | -0.46 (0.28) | -1.52*** (0.31) | -1.25*** (0.27) | -0.82*** (0.22) | -0.76*** (0.21) | | |
| Sub-Sample | 0.40 (0.32) | 0.42 (0.30) | 0.22 (0.29) | 0.39 (0.27) | 0.04 (0.27) | -0.04 (0.25) | -0.43 (0.29) | -0.24 (0.28) | 0.04 (0.22) | 0.01 (0.21) | | |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Constant | 3.69*** (0.40) | 3.69*** (0.38) | 3.69*** (0.38) | 3.47*** (0.36) | 3.84*** (0.39) | 3.73*** (0.36) | 4.09*** (0.39) | 3.93*** (0.36) | 3.46*** (0.31) | 3.38*** (0.29) | | |
| Observations | 972 | 1,201 | 1,102 | 1,393 | 1,195 | 1,544 | 1,107 | 1,382 | 1,177 | 2,275 | | |

Note: *p<0.05; **p<0.01; ***p<0.001

As an additional test, we model the correlates of holding benevolent and hostile sexist beliefs. In accordance with Winter's (2018) findings, Republicans in our sample are more likely than Democrats to be hostile (though not benevolent) sexists. This suggests that there may be a connection between party identification and first-order beliefs in gender stereotypes. Specifically, the correlation between our measure of hostile sexism and a binary measure of whether a respondent is a Republican is $\rho = 0.46$, while the correlation is $\rho = 0.02$ for benevolent sexism. In Table A.9, we show that the relationship between hostile sexism and party identification is also statistically significant in a model with controls, while the relationship between benevolent sexism and party identification is not. Interestingly, older respondents and female respondents are less likely to hold hostile sexist views, but more likely to hold benevolent sexist views.

Table A.9: Correlates of Sexism

| | Hostile Sexism | Benevolent Sexism |
|------------------------|---------------------|---------------------|
| | (1) | (2) |
| Republican | 0.19*** (0.01) | 0.001 (0.01) |
| Militant Assertiveness | 0.17*** (0.02) | 0.04** (0.02) |
| Age | -0.03*** (0.004) | 0.02*** (0.003) |
| Female | -0.09*** (0.01) | 0.02*** (0.01) |
| Education | -0.03*** (0.004) | -0.02*** (0.004) |
| Income | -0.01*** (0.003) | -0.01*** (0.003) |
| Party US President | -0.01 (0.01) | -0.002 (0.01) |
| Gender US President | 0.01 (0.01) | -0.003 (0.01) |
| Gender Opponent | 0.02* (0.01) | -0.01 (0.01) |
| Engage | 0.01 (0.01) | 0.01 (0.01) |
| Not Engage | 0.01 (0.01) | 0.01 (0.01) |
| Sexism Order | 0.01 (0.01) | -0.01* (0.01) |
| Constant | 0.45*** (0.02) | 0.57*** (0.02) |
| Observations | 2,325 | 2,324 |

Note:

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

mTurk Experiment

Survey Text

Sexism (Question order randomized; half of respondents received these questions pre-treatment and half post-treatment):

1. Please select the extent to which you agree or disagree with the following statements:

| | Strongly Disagree | Somewhat Disagree | Slightly Disagree | Slightly Agree | Somewhat Agree | Strongly Agree |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| In a disaster, there is no need for women to be rescued by men. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Many women have a quality of purity that few men possess. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Feminists are making reasonable demands of men. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Women who complain about discrimination often cause more problems than they solve. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Foreign Policy Dispositions Pre-Test (Question order randomized):

2. Please select the extent to which you agree or disagree with the following statements:

| | Strongly Disagree | Somewhat Disagree | Neither Agree Nor Disagree | Somewhat Agree | Strongly Agree |
|--|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|
| The best way to ensure world peace is through American military strength. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Going to war is unfortunate, but sometimes the only solution to international problems. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| The use of military force only makes problems worse. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

3. Please select the statement that best reflects your opinion:

Vastly Very Not so Not at all
Superior Superior Superior Superior

**How superior is the
United States compared
to other nations?**

4. Please select the statement that best reflects your opinion:

The United States can trust other nations.

**The United States cannot be too careful
in dealing with other nations.**

5. Please select the statement that best reflects your opinion:

Very Not
Many Many Many None

**How many things about
America make you ashamed?**

Random Assignment to Scenario:

The following questions are about U.S. relations with other countries around the world. You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach U.S. leaders could take in the future and ask whether you approve or disapprove.

Treatment Screen: [1 of 24 scenarios—equal assignment probability; names randomized; quota implemented based on party identification to ensure an equal number of Republicans and Democrats in the survey]

Stay Out Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] U.S. President, [Erica/Eric, Stephanie/Steven] Smith, says the United States will stay out of the conflict. The attacking country continues to invade. In the end, the US president decides not to send troops, and the attacking country ultimately gains 20% of the contested territory.

Not Engage Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] US President, [Erica/Eric, Stephanie/Steven] Smith, says that if the attack continues, the US military will push out the invaders. The attacking country continues to invade. In the end, the US president does not send troops, and the attacking country ultimately gains 20% of the contested territory.

Engage Condition: A country sends its military to take over a neighboring country. The attacking country is controlled by a [female/male] leader.

The [Republican/Democratic] US President, [Erica/Eric, Stephanie/Steven] Smith, says that if the attack continues, the US military will push out the invaders. The attacking country continues to invade. In the end, the US president orders the US military to engage. Ultimately, the attacking country gains 20% of the contested territory and the US experiences zero casualties.

To summarize:

Stay Out

- The attacking country is led by a [female/male] leader
- The U.S. President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith
- (Erica/Eric, Stephanie/Steven) Smith says the United States will stay out of the conflict
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does not send troops
- The attacking country ultimately gains 20% of the contested territory

Not Engage

- The attacking country is led by a [female/male] leader
- The U.S. President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith
- (Erica/Eric, Stephanie/Steven) Smith says that if the attack continues, the U.S. military will push out the invaders
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does not send troops
- The attacking country ultimately gains 20% of the contested territory

Engage

- The attacking country is led by a [female/male] leader
- The U.S. President is a [Republican/Democrat] named [Erica/Eric, Stephanie/Steven] Smith
- (Erica/Eric, Stephanie/Steven) Smith says that if the attack continues, the U.S. military will push out the invaders
- The attacking country continues to invade and [Erica/Eric, Stephanie/Steven] Smith does decide to send troops
- The attacking country ultimately gains 20% of the contested territory
- The US experiences zero casualties

Dependent Variable:

6. What are your views about how the US President handled the situation?

| | | | | | | |
|------------------------|-----------------------|-----------------------------|-----------------------------------|--------------------------|-----------------------|-----------------------|
| Strongly Disapprove | Disapprove | Lean Toward Disapproving | Neither Approve Nor Disapprove | Lean Toward Approving | Approve | Strongly Approve |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Placebo & Manipulation Check Questions:

8. What is your best estimate of how democratic the attacking country is, on a scale of -10 to +10, where -10 is the least democratic and +10 is the most democratic? Next to the options are some example countries to help calibrate your answer.

| | | | | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| -10 (North Korea) | -8 | -6 (China) | -4 | -2 (Jordan) | 0 | 2 (Algeria) | 4 | 6 (Pakistan) | 8 | 10 (Canada) |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

9. What is your best guess of what race the US President is?

| | | | | |
|-----------------------|----------------------------|-----------------------|-----------------------|-----------------------|
| Caucasian/ White | African American/ Black | Asian | Hispanic/ Latino | Other |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

10. What was the name of the US President?

Eric Erica Steven Stephanie None of
the Above

11. What was the sex of the leader of the attacking country?

Male Female Not Stated

mTurk Results

Table A.10 presents average treatment effects calculated from our pilot mTurk study. As hypothesized, inconsistency costs are significantly greater for the FM, FF, and MF dyads compared to the MM baseline. Though the results for belligerence costs are also in the expected direction, they are not statistically significant.

Table A.10: Difference in Mean Disapproval on a 7-Point Scale Compared to the Male-Male Baseline (mTurk)

| Dyad (Baseline = Male-Male) | Audience Cost (1) | Inconsistency Cost (2) | Belligerence Cost (3) |
|--------------------------------|-------------------------|------------------------------|-----------------------------|
| Female-Male | +0.08 (1.13 – 1.05) | +0.43* (2.14 – 1.71) | -0.35 (-1.01 – -0.66) |
| Female-Female | +0.31 (1.37 – 1.05) | +0.40* (2.11 – 1.71) | -0.09 (-0.75 – -0.66) |
| Male-Female | +0.42* (1.48 – 1.05) | +0.46** (2.17 – 1.71) | -0.04 (-0.69 – -0.66) |

Note: Results depict average treatment effects (ATE) for a 7-point measure of disapproval calculated from 2,000 bootstraps. The main quantities reflect the average difference in disapproval on a 7-point scale for the dyad compared to the male-male baseline. For example, average disapproval is 0.43 points higher on a 7-point scale for a female president that acts inconsistently against a foreign male leader compared to a male president that acts inconsistently against a foreign male leader. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$. The mean disapproval for the two experimental groups used to calculate ATE are in parentheses.

References for Supplementary Materials

1. Bauer, Nichole M. 2017. "The Effects of Counterstereotypic Gender Strategies on Candidate Evaluations." *Political Psychology* 38 (2), 279-95.
2. Christensen, Martin Iversen. 2015. "Worldwide Guide to Women in Leadership." Accessed January 15, 2020. <http://www.guide2womenleaders.com/index.html>.
3. Ellemers, Naomi. 2018. "Gender Stereotypes." *Annual Review of Psychology* 69, 275-98.
4. Glick, Peter, and Susan T. Fiske. 1996. "The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism." *Journal of Personality and Social Psychology* 70(3), 491-512.
5. Goemans, Henk E., Kristian Skrede Gleditsch, and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research* 46(2), 269-83.
6. Hayes, Danny. 2011. "When Gender and Party Collide: Stereotyping in Candidate Trait Attribution." *Politics & Gender* 7(2), 133-65.
7. Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. 2015. *Why Leaders Fight*. New York, NY.: Cambridge University Press.
8. Jalalzai, Farida. 2013. *Shattered, Cracked, or Firmly Intact? Women and the Executive Glass Ceiling Worldwide*. New York, NY: Oxford University Press.
9. 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), 234-49.
10. Post, Abigail S., and Paromita Sen. 2020. "Why Can't a Woman Be More Like a Man? Female Leaders in Crisis Bargaining." *International Interactions* 46(1), 1-27.
11. Sanbonmatsu, Kira, and Kathleen Dolan. 2009. "Do Gender Stereotypes Transcend Party?" *Political Research Quarterly* 62(3), 485-94.
12. Winter, Nicholas J.G. 2018. "Ambivalent Sexism and Election 2016." Unpublished Manuscript, University of Virginia.