**Improving the Measurement of Hostile Sexism**

|  |  |
| --- | --- |
| Allison M. N. Archer[[1]](#footnote-1)\*University of Houston3551 Cullen Boulevard, Room 436Houston, TX 77204-3011713-743-2258amarcher3@uh.edu | Scott CliffordUniversity of Houston3551 Cullen Boulevard, Room 424Houston, TX 77204-3011713-743-2538sclifford@uh.edu |

**Supplementary Material for “Improving the Measurement of Hostile Sexism”**

**Table of Contents**

Page 1. Question Wording of the AD and IS Measures of Hostile Sexism

Page 2-6. Question Wording for Study 1

Page 7-9. Question Wording for Study 2

Page 10-11. Question Wording for Study 3

Page 12. Table S1. Sample Demographics

Page 13. Figure S1. Distributions of the 5-Item and 11-Item Hostile Sexism Scales (Study 2)

Page 14-16. Estimation of a Latent Measure of MeToo Attitudes (Studies 2 & 3)

Page 17-18. Analysis of Reversed Items in the 11-Item Scales (Study 2)

Page 18. Table S2. The Impact of Reversed Items

Page 19. Discriminant Validity in the 11-Item Scales (Study 2)

Page 20. Table S3. Feeling Thermometers by Candidate Gender and Partisanship (Study 2)

Page 21-22. Alternative Analyses of the Feeling Thermometer Outcomes (Study 2)

Page 22. Table S4. Feeling Thermometers: Gender Differences by In- and Out-Party (Study 2)

Page 23. Table S5. Predictive Validity for the 11-Item Hostile Sexism Scales (Study 2)

Page 24. Figure S2. Predictive Validity for Individual Politicians (Study 2)

Page 25. Table S6. Analysis of Perceptions of Sexual Misconduct Allegations (Study 2)

Page 26. Question Wording of the Benevolent and Modern Sexism Scales

**Question Wording of the AD and IS Measures of Hostile Sexism**

|  |  |  |
| --- | --- | --- |
| AD Question | IS Question | IS Response Options |
| Women seek to gain power by getting control over men (1) | How often do women seek to gain power by getting control over men? | Never, rarely, sometimes, often, always |
| Women exaggerate problems they have at work (2) | How often do women exaggerate problems they have at work? | Never, rarely, sometimes, often, always |
| Once a woman gets a man to commit, she puts him on a tight leash (3) | When a woman gets a man to commit, how often does she put him on a tight leash? | Never, rarely, sometimes, often, always |
| Women are too easily offended (4) | Would you say that women are offended too often, or not offended often enough? | Women are offended [much too often, a bit too often, about the right amount, not quite often enough, not nearly often enough] |
| Most women fail to appreciate fully all that men do for them (5) | Do you think women give men too much credit or too little credit for what men do for them?  | Women give men [far too much, a bit too much, about the right amount of, a bit too little, far too little] credit |
| Most women interpret innocent remarks as being sexist (6) | How often do women interpret innocent remarks as being sexist? | Never, rarely, sometimes, often, always |
| When women lose to men in a fair competition, they typically complain about being discriminated against (7) | When women lose to men in a fair competition, how often do they complain about being discriminated against? | Never, rarely, sometimes, often, always |
| Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for “equality.” (8) | How many women use the guise of “equality” to ask for special favors, such as hiring policies, that favor them over men? | None, very few, some, most, all |
| Feminists are making entirely reasonable demands of men (9) | Are the demands that feminists make of men reasonable or not? | Not reasonable at all, not too reasonable, somewhat reasonable, mostly reasonable, completely reasonable |
| Feminists are not seeking for women to have more power than men (10) | How many feminists are actually just seeking to have more power than men? | None, very few, some, most, all |
| There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances (11) | How many women get a kick out of teasing men by seeming sexually available and then refusing male advances? | None, very few, some, most, all |

**Question Wording for Study 1**

*The 5-item versions of the IS Hostile Sexism scale and AD Hostile Sexism scale were displayed in randomized order. See page 1 for exact wording. The IS Hostile Sexism scale displayed questions vertically and individually. The AD Hostile Sexism scale displayed questions horizontally and in a grid.*

*Respondents were randomly assigned to one of the following pairs at the end of Study 1. Answer options for this section were presented in randomized order.*

[self\_credit] On the previous two pages, you answered two different types of questions on the same topic. Below are two different versions of the same question. Which version do you think better measured your own opinion on the topic?

1--

2--

[self\_leash] On the previous two pages, you answered two different types of questions on the same topic. Below are two different versions of the same question. Which version do you think better measured your own opinion on the topic?

1--

2--

[self\_offend] On the previous two pages, you answered two different types of questions on the same topic. Below are two different versions of the same question. Which version do you think better measured your own opinion on the topic?

1--

2--

[self\_power] On the previous two pages, you answered two different types of questions on the same topic. Below are two different versions of the same question. Which version do you think better measured your own opinion on the topic?

1--

2--

[self\_problems] On the previous two pages, you answered two different types of questions on the same topic. Below are two different versions of the same question. Which version do you think better measured your own opinion on the topic?

1--

2--

**Question Wording for Study 2**

*Attention Check*

People are very busy these days and many do not have time to follow what goes on in the government. We are testing whether people read questions. To show that you've read this much, please answer "Not interested at all".

Not interested at all (1)

Not too interested (2)

Somewhat interested (3)

Very interested (4)

Extremely interested (5)

*Feeling Thermometers*

Next, we would like to get your opinion on several groups, organizations, and individuals in our society. For each group, organization, or individual, we will ask you to tell us how you feel about the group, organization or individual on a feeling thermometer using a scale of 0 to 100. The higher the number, the warmer or more favorable you feel toward that group, organization, or individual; the lower the number, the colder or less favorable you feel. You can pick any number between 0 and 100.

Kamala Harris

Hillary Clinton

Donald Trump

Sarah Palin

Bernie Sanders

Elizabeth Warren

Barack Obama

Joe Biden

Amy Coney Barrett

Mike Pence

*Views on MeToo & Workplace Sexual Harassment Training (questions in this section [except* backlash*]* *adapted from Archer & Kam 2020)*

[movement] The #MeToo movement aims to end sexual violence and harassment; it rose to prominence in fall 2017. In your opinion, has the #MeToo movement gone too far or not far enough? Or is the level of activity about right?

 1—Has gone too far

 2—Is about right

 3—Has not gone far enough

 4—Not sure

[responsibility] Do you regard sexual harassment in the workplace as the responsibility of the company to prevent or solve, or the responsibility of the individuals involved to prevent or solve?

 1—Entirely the responsibility of the company to prevent or solve

2—Mostly the responsibility of the company to prevent or solve

 3—Equally the responsibility of the company and the individuals involved

4—Mostly the responsibility of the individuals involved to prevent or solve

 5—Entirely the responsibility of the individuals involved to prevent or solve

[training] Do you support or oppose corporations requiring employees to undergo mandatory sexual harassment training?

1 – Strongly support

2 – Somewhat support

3 – Neither support nor oppose

4 – Somewhat oppose

5 – Strongly oppose

[backlash] There has been some discussion of backlash to the #MeToo movement in the workplace. Which of the following below do you think have happened as a consequence of the #MeToo movement? (Check all that apply.)

1—Employers have been more reluctant to hire attractive women

2—Employers have been less likely to hire women for jobs that require close interpersonal interactions with men (e.g., travel)

3—Men have been more likely to exclude women from social interactions

4—Men have been more reluctant to have one-on-one meetings with women

5—Men have been more hesitant to mentor women

6—None of the above

[isolated] As you may know, over the past few years, many high-profile individuals have been accused of sexual misconduct to varying extents. Which comes closer to your view about these recent allegations in the news?

 1—They are mainly isolated incidents of individual misconduct

 2—They mainly reflect widespread problems in society

*Credibility of Accusations*

How credible or not credible do you find the sexual misconduct allegations against each of the following individuals?

Matt Lauer

Harvey Weinstein

Bill Cosby

Kevin Spacey

1 – Not credible at all

2 – Not too credible

3 – Somewhat credible

4 – Very credible

5 – Completely credible

6 – Not sure

*Random assignment to one of the Hostile Sexism Scale Formats:*

*Hostile Sexism – Item-Specific Format (items displayed individually in randomized order with response options displayed vertically.) See page 1 for exact wording.*

*Hostile Sexism – Agree-Disagree Format (items displayed in a grid in randomized order.) See page 1 for exact wording.*

*Feelings of Control Scale*

{Control 1}

How strongly do you agree or disagree with the following statements?

The really good things that happen to me are mostly luck.

There's no sense planning a lot--if something good is going to happen it will.

Most of my problems are due to bad breaks.

I have little control over the bad things that happen to me.

1 – Agree strongly

2 – Agree somewhat

3 – Agree slightly

4 – Disagree slightly

5 – Disagree somewhat

6 – Disagree strongly

{Control 2}

How strongly do you agree or disagree with the following statements?

I am responsible for my own success.

I can do just about anything I really set my mind to.

My misfortunes are the result of mistakes I have made.

I am responsible for my failures.

Select “Disagree somewhat” to show you’re reading the questions. {*This item is an IMC not included in the analysis of feelings of control.*}

1 – Agree strongly

2 – Agree somewhat

3 – Agree slightly

4 – Disagree slightly

5 – Disagree somewhat

6 – Disagree strongly

**Question Wording for Study 3**

*Attention Check*

People are very busy these days and many do not have time to follow what goes on in the government. We are testing whether people read questions. To show that you've read this much, please answer "Not interested at all".

Not interested at all (1)

Not too interested (2)

Somewhat interested (3)

Very interested (4)

Extremely interested (5)

*Note: question wording for the feeling thermometers and #MeToo outcomes was identical to Study 2. See above for details. The only differences are that the Backlash question and Credibility of Accusations questions were not asked here.*

*Hostile Sexism – Item-Specific Format (items displayed individually in randomized order with response options displayed vertically.) See page 1 for exact wording.*

*Hostile Sexism – Agree-Disagree Format (items displayed individually in randomized order with response options displayed vertically.)*

How strongly do you agree or disagree with the following statements?

Women seek to gain power by getting control over men (1)

Women exaggerate problems they have at work (2)

Once a woman gets a man to commit, she puts him on a tight leash (3)

Women are too easily offended (4)

Most women fail to appreciate fully all that men do for them (5)

Most women interpret innocent remarks as being sexist (6)

When women lose to men in a fair competition, they typically complain about being discriminated against (7)

Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for “equality.” (8)

Feminists are making entirely reasonable demands of men (9)

Feminists are not seeking for women to have more power than men (10)

There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances (11)

1 – Agree strongly

2 – Agree somewhat

3 – Neither agree nor disagree

4 – Disagree somewhat

5 – Disagree strongly

*Feelings of Control Scale (items displayed individually and in randomized order with response options displayed vertically)*

How strongly do you agree or disagree with the following statements?

The really good things that happen to me are mostly luck.

There's no sense planning a lot--if something good is going to happen it will.

Most of my problems are due to bad breaks.

I have little control over the bad things that happen to me.

I am responsible for my own success.

I can do just about anything I really set my mind to.

My misfortunes are the result of mistakes I have made.

I am responsible for my failures.

1 – Agree strongly

2 – Agree somewhat

3 – Neither agree nor disagree

4 – Disagree somewhat

5 – Disagree strongly

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   |   |   |   |   |  |
|   | **Table S1. Sample Demographics** |  |
|   |   |   | Study 1 | Study 2 | Study 3 |
|   |  |  | MTurk | Lucid | MTurk |
|   | Partisan Identity |   |   |  |   |
|   |   | Democrat | 52% | 39% | 46% |
|   |   | Republican | 20% | 29% | 23% |
|   |   | Independent | 28% | 32% | 31% |
|   | Ideology |   |   |  |   |
|   |   | Liberal | - | 33% | 53% |
|   |   | Conservative | - | 33% | 27% |
|   |   | Moderate | - | 34% | 20% |
|  | College Degree |  | 54% | 44% | 60% |
|   | Race and Ethnicity |   |   |  |   |
|   |   | White | 79% | 74% | 74% |
|   |   | Black | 6% | 10% | 9% |
|   |   | Asian | 8% | 6% | 8% |
|   |   | Hispanic | 5% | 7% | 6% |
|   |   | Other | 2% | 3% | 6% |
|   | Median Age  | 36 | 45 | 38 |  |
|   | Male | 62% | 56% | 49% |  |
|   |   |   |   |   |  |

Note: MTurk and Lucid Theorem provide convenience samples. Respondents in Studies 1 and 3 were paid $0.50 for their participation via MTurk; respondents in Study 2 were compensated by Lucid Theorem. Though we did not record device type in Studies 1 and 2, we did use Qualtrics software for both surveys. Qualtrics automatically adjusts the format of response scales to better fit the device, and grids in particular are adjusted so that they are displayed much more like single items. Despite this, in Study 3, we recorded information regarding OS, browser and screen resolution. From this information, we can infer that only 103 of our 2,003 respondents in that study used a smartphone.

**Figure S1. Distributions of the 5-Item and 11-Item Hostile Sexism Scales (Study 2)**



**Estimation of a Latent Measure of MeToo Attitudes (Studies 2 & 3)**

We measured attitudes toward the #MeToo movement with a series of five questions. We exclude one of these questions from our analysis because it was nearly unrelated to the rest of the items and thus clearly did not tap into the same latent variable. Of the remaining four items, one is dichotomous and the other three are ordinal. Thus, we used a hybrid item response model, which combines a graded response model and a 2-parameter logistic model, to estimate a single latent dimension of #MeToo attitudes. The discrimination and difficulty parameters are shown in the tables below by study. The distributions of the latent variable are shown in the figures below also by study.

**Study 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item |   | Movement | Responsibility | Training | Isolated |
| Discrimination |  | 1.16 | 0.69 | 1.42 | 1.14 |
|  |  | (.10) | (.07) | (.13) | (.11) |
|   |   |   |   |   |   |
| Difficulty | 2 | -1.47 | -2.14 | -0.15 | 0.45 |
|  |  | (.10) | (.20) | (.04) | (.05) |
|  | 3 | 1.11 | -0.45 | 0.96 |  |
|  |  | (.08) | (.07) | (.07) |  |
|  | 4 |  | 3.35 | 2.12 |  |
|  |  |  | (.30) | (.13) |  |
|  | 5 |  | 5.10 | 2.72 |  |
|   |   |   | (.47) | (.17) |   |



**Study 3**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item |   | Movement | Responsibility | Training | Isolated |
| Discrimination |  | 1.83 | 0.86 | 1.85 | 2.33 |
|  |  | (.12) | (.06) | (.12) | (.20) |
|   |   |   |   |   |   |
| Difficulty | 2 | -0.99 | -2.68 | -0.03 | 0.61 |
|  |  | (.05) | (.19) | (.04) | (.04) |
|  | 3 | 0.97 | -0.66 | 1.05 |  |
|  |  | (.05) | (.07) | (.05) |  |
|  | 4 |  | 2.71 | 1.82 |  |
|  |  |  | (.19) | (.08) |  |
|  | 5 |  | 5.00 | 2.49 |  |
|   |   |   | (.38) | (.12) |   |



**Analysis of Reversed Items in the 11-Item Scales (Study 2)**

Both short versions formed reliable scales, though the AD version had higher reliability (α = .92) than the IS version (α = .71). While this may seem like a virtue of the original scale, it is also consistent with the possibility that the AD format inflates the relationship between items. As expected, the full version of the IS scale was more reliable than its short version (α = .87). However, for the AD scale, the full scale was actually less reliable than the short scale (α = .89). This may be because the full scale includes reversed items and method bias pushed respondents in the opposite direction on these items.

We now examine how reversed items affect scale coherence. Again, such items should be more problematic in the AD scale than the IS because of the response format. The AD scale contains three reversed items (9, 10, 11), while the IS scale contains two reversed items (4, 9). To test the impact of reversal, we created a subset of each scale that consists of only the items that are forward-worded in *both* versions of the scale (1-3, 5-8). We then separately analyzed the correlation between this subscale and each of the remaining items in the scale. The results are shown in Table S2. We start with item 4, which was reversed in the IS scale, but *not* in the AD scale. Due to the reversal, we should expect a stronger correlation between this item and the forward-worded items in the AD scale than in the IS scale. As expected, while the correlation within the AD scale is high (*r* = .81), it is only moderate in the IS scale (*r* = .43). Thus, even in the IS scale, reversals seem to hurt internal coherence. However, we expect that this effect should be larger for the AD scale. The next item, 9, was reversed in *both* scales. Consistent with expectations, this item fares better in the IS scale (*r* = .31) than in the AD scale (*r* = .15). The next two items are reversed in the AD scale, but not the IS scale. Here, we find dramatic differences in the correlations. While these items are strongly related to the rest of the scale in the IS version (*r* = .69, *r* = .67), they are mostly unrelated to the remaining items in the AD scale (*r* = .21, *r* = -.04). Overall, it’s clear that reversed items are less strongly related to other items in the scale for both the IS and AD hostile sexism scales. However, it’s also clear that the impact of reversals is consistently smaller for the IS version than for the AD.

**Table S2. The Impact of Reversed Items**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Direction | Original (AD) | Revised (IS) | Difference |
| 4 | Revised Reversed | 0.81 | 0.43 | -0.39 |
| 9 | Both Reversed | 0.15 | 0.31 | 0.16 |
| 10 | Original Reversed | 0.21 | 0.69 | 0.48 |
| 11 | Original Reversed | -0.04 | 0.67 | 0.71 |

**Discriminant Validity in the 11-Item Scales (Study 2)**

We also conducted an analysis of discriminant validity for the full 11-item versions of each scale. The results are quite similar, though slightly less dramatic. Starting with the AD scale, there is a large difference in the correlations between the two halves of the feelings of control scale (forward-worded: *r* = .33; reversed: *r* = -.02). But the IS scale still showed a slightly smaller gap between the two scales (forward: *r* = .20; reverse: *r* = -.10). We conducted the same bootstrapping procedure described in the main text, and the difference in correlations was larger for the AD scale in 81.1% of the 1,000 resamples. Thus, the results for the full scale are less clearly in favor of the IS battery, likely due to the inclusion of the reversed items in the full AD scale. While this may seem like a solution, recall that the reversed items in the AD scale are largely uncorrelated with the remaining items in the scale.

**Table S3. Feeling Thermometers by Candidate Gender and Partisanship (Study 2)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Democratic Women |  | Republican Men |  | Democratic Men |  | Republican Women |  |
|  |  |  |  |  |  |  |  |  |
|  | Beta (SE) |  *p* | Beta (SE) | *p* | Beta (SE) | *p* | Beta (SE) | p |
| Hostile Sexism (HS) | -8.676\*\*\* | 0.001 | 25.403\*\*\* | 0.000 | -12.859\*\*\* | 0.000 | 19.387\*\*\* | 0.000 |
|  | (2.530) |  | (2.917) |  | (2.547) |  | (2.830) |  |
| IS Format (IS) | 6.098\* | 0.022 | -3.904 | 0.209 | -0.620 | 0.817 | 2.834 | 0.345 |
|  | (2.669) |  | (3.107) |  | (2.681) |  | (2.999) |  |
| IS x HS | -8.115 | 0.101 | 3.218 | 0.572 | 4.322 | 0.385 | -6.710 | 0.224 |
|  | (4.940) |  | (5.696) |  | (4.969) |  | (5.515) |  |
| Partisan Identity | -7.338\*\*\* | 0.000 | 8.508\*\*\* | 0.000 | -8.429\*\*\* | 0.000 | 3.507\*\*\* | 0.000 |
|  | (0.289) |  | (0.331) |  | (0.292) |  | (0.320) |  |
| Ideology | -3.976\*\*\* | 0.000 | 3.886\*\*\* | 0.000 | -3.743\*\*\* | 0.000 | 3.554\*\*\* | 0.000 |
|  | (0.367) |  | (0.424) |  | (0.372) |  | (0.409) |  |
| Political Interest | 4.367\*\*\* | 0.000 | 2.698\*\*\* | 0.000 | 3.232\*\*\* | 0.000 | 3.176\*\*\* | 0.000 |
|  | (0.446) |  | (0.511) |  | (0.452) |  | (0.499) |  |
| Education | 1.266\*\*\* | 0.000 | 0.310 | 0.444 | 0.516 | 0.145 | 0.895\* | 0.023 |
|  | (0.351) |  | (0.405) |  | (0.354) |  | (0.392) |  |
| Age | 0.034 | 0.299 | -0.081\* | 0.030 | -0.107\*\*\* | 0.001 | 0.036 | 0.315 |
|  | (0.032) |  | (0.037) |  | (0.032) |  | (0.036) |  |
| Male | 0.103 | 0.922 | -1.540 | 0.199 | -0.433 | 0.680 | 1.686 | 0.147 |
|  | (1.046) |  | (1.199) |  | (1.050) |  | (1.162) |  |
| Asian | -3.730 | 0.141 | 2.292 | 0.442 | -3.490 | 0.170 | 4.169 | 0.148 |
|  | (2.530) |  | (2.984) |  | (2.542) |  | (2.883) |  |
| Hispanic | -2.124 | 0.395 | -0.584 | 0.843 | -1.448 | 0.562 | 1.864 | 0.514 |
|  | (2.497) |  | (2.950) |  | (2.496) |  | (2.856) |  |
| Native American | -10.723\* | 0.036 | 3.514 | 0.539 | -8.931 | 0.074 | 2.425 | 0.660 |
|  | (5.116) |  | (5.715) |  | (4.993) |  | (5.510) |  |
| White | -7.936\*\*\* | 0.000 | 5.791\*\* | 0.005 | -8.226\*\*\* | 0.000 | 2.199 | 0.266 |
|  | (1.741) |  | (2.057) |  | (1.734) |  | (1.978) |  |
| Other Race | -9.505\* | 0.010 | -1.801 | 0.678 | -4.812 | 0.187 | -3.084 | 0.457 |
|  | (3.691) |  | (4.340) |  | (3.644) |  | (4.149) |  |
| Constant | 73.184\*\*\* | 0.000 | -28.123\*\*\* | 0.000 | 102.636\*\*\* | 0.000 | -18.711\*\*\* | 0.000 |
|  | (3.029) |  | (3.517) |  | (3.026) |  | (3.406) |  |
| *R*2 | 0.54 |  | 0.51 |  | 0.57 |  | 0.26 |  |
| *N* | 2130 |  | 2163 |  | 2187 |  | 2142 |  |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Alternative Analyses of the Feeling Thermometer Outcomes (Study 2)**

Given the surprisingly poor performance of the AD scale in predicting gendered candidate evaluations, we sought to dig deeper into this finding. One possible explanation stems from the fact that the scale seems to respond to two signals – a candidate’s gender and their partisanship. To investigate this possibility, we created four average ratings: Democratic women, Democratic men, Republican women, and Republican men.[[2]](#footnote-2) Previously, partisanship was held constant in the analyses that differenced feeling thermometers by gender, but this disaggregation allows us to better understand evaluations by PID *and* gender. We use the same modeling approach as in the main text to predict each of the four outcomes. In none of the four tests are the coefficients for the two scales significantly different from each other, but the patterns are suggestive.

We start with cases in which the signals are aligned – Democratic women and Republican men. Both versions of the scale significantly predict more negative evaluations of Democratic women. However, the coefficient for the IS format is nearly twice as large (*b* = -16.8, *p* < .001) as the coefficient for the AD format (*b* = -8.7, *p* < .001). Both scales also strongly predict more positive evaluations of Republican men (IS: *b* = 28.6, *p* < .001; AD: *b* = 25.4, *p* < .001).

 We now turn to cases in which the signals conflict – Republican women and Democratic men. Both scales predict more *positive* views of Republican women, though the effect is somewhat more positive for the AD scale (*b* = 19.4, *p* < .001) than the IS scale (*b* = 12.7, *p* = .010). As for Democratic men, both scales predict more *negative* evaluations, though the effect is again slightly stronger for the AD scale (*b* = -12.9, *p* < .001) than the IS scale (*b* = -8.5, *p* = .056). These two findings suggest that hostile sexism responds more to candidate partisanship than candidate gender, perhaps especially so for the AD scale.

As an alternative strategy to analyzing the feeling thermometer data, we created two new dependent variables based on whether the candidates were from the respondent’s own party or the opposing party. Pure independents were excluded from this analysis. In the first column, the dependent variable is the average rating of in-party men minus the average rating of in-party women. The second column shows the same calculation but for out-party candidates. Starting with in-party candidates, there is a significant interaction between hostile sexism and the scale format (*p* = .001). Specifically, while higher values of sexism on the IS scale predict a greater preference for men (*b* = 16.2, *p* < .001), it does not do so with the AD scale (*b* = -0.4, *p* = .890). Thus, this analysis again suggests the IS format has better predictive power. Turning to the out-party measure, we find little evidence that either the IS (*b* = 3.1, *p* = .434) or the AD (*b* = 1.0, *p* = .656) measure predicts preference for men. This is likely because respondents are unlikely to like out-party candidates, regardless of their gender.

**Table S4. Feeling Thermometers: Gender Differences by In- and Out-Party (Study 2)**

|  |  |  |
| --- | --- | --- |
|  | In-Party Candidates | Out-Party Candidates |
|  |  |  |  |  |
|  | Beta (SE) | *p* | Beta (SE) | *p* |
| Hostile Sexism | -0.354 | 0.890 | 1.013 | 0.656 |
|  | (2.556) |  | (2.275) |  |
| IS Format | -7.995\*\* | 0.002 | -2.808 | 0.241 |
|  | (2.617) |  | (2.392) |  |
| IS Format x Hostile Sexism | 16.542\*\*\* | 0.001 | 2.107 | 0.634 |
|  | (4.847) |  | (4.429) |  |
| Partisan Identity | 1.165\*\*\* | 0.000 | 2.728\*\*\* | 0.000 |
|  | (0.268) |  | (0.237) |  |
| Ideology | 0.636 | 0.070 | -0.049 | 0.874 |
|  | (0.351) |  | (0.311) |  |
| Political Interest | -1.966\*\*\* | 0.000 | 0.175 | 0.674 |
|  | (0.467) |  | (0.417) |  |
| Education | -1.386\*\*\* | 0.000 | 0.161 | 0.607 |
|  | (0.350) |  | (0.313) |  |
| Age | -0.198\*\*\* | 0.000 | -0.064\* | 0.025 |
|  | (0.031) |  | (0.029) |  |
| Male | -3.405\*\* | 0.001 | -0.523 | 0.576 |
|  | (1.039) |  | (0.935) |  |
| Asian | -0.689 | 0.787 | -1.169 | 0.616 |
|  | (2.554) |  | (2.329) |  |
| Hispanic | -1.556 | 0.542 | 0.374 | 0.873 |
|  | (2.549) |  | (2.335) |  |
| Native American | 3.601 | 0.522 | 1.006 | 0.848 |
|  | (5.621) |  | (5.242) |  |
| White | 2.646 | 0.128 | -0.311 | 0.842 |
|  | (1.736) |  | (1.557) |  |
| Other Race | -1.510 | 0.707 | 7.138 | 0.064 |
|  | (4.012) |  | (3.847) |  |
| Constant | 30.797\*\*\* | 0.000 | -9.859\*\*\* | 0.000 |
|  | (3.131) |  | (2.826) |  |
| *R*2 | 0.105 |  | 0.132 |  |
| *N* | 1780 |  | 1656 |  |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table S5. Predictive Validity for the 11-Item Hostile Sexism Scales (Study 2)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | #MeToo Attitudes |  | Feeling Thermometers (Men - Women) |  |
|  |  |  |  |  |
|  | Beta (SE) | *p* | Beta (SE) | *p* |
| Hostile Sexism (11-item) | 1.241\*\*\* | 0.000 | 1.890 | 0.311 |
|  | (0.088) |  | (1.865) |  |
| IS Format | -0.190\*\* | 0.010 | -4.095\*\* | 0.009 |
|  | (0.073) |  | (1.573) |  |
| IS Format x Hostile Sexism (11) | 0.225 | 0.115 | 6.545\* | 0.032 |
|  | (0.143) |  | (3.058) |  |
| Partisan Identity | 0.043\*\*\* | 0.000 | 1.367\*\*\* | 0.000 |
|  | (0.008) |  | (0.162) |  |
| Ideology | 0.065\*\*\* | 0.000 | 0.220 | 0.288 |
|  | (0.010) |  | (0.207) |  |
| Political Interest | -0.056\*\*\* | 0.000 | -0.701\*\* | 0.005 |
|  | (0.012) |  | (0.249) |  |
| Education | 0.003 | 0.771 | -0.528\*\* | 0.007 |
|  | (0.009) |  | (0.195) |  |
| Age | 0.005\*\*\* | 0.000 | -0.134\*\*\* | 0.000 |
|  | (0.001) |  | (0.018) |  |
| Male | 0.161\*\*\* | 0.000 | -1.858\*\* | 0.001 |
|  | (0.028) |  | (0.582) |  |
| Asian | 0.037 | 0.577 | -0.394 | 0.782 |
|  | (0.067) |  | (1.424) |  |
| Hispanic | -0.069 | 0.293 | 0.031 | 0.983 |
|  | (0.065) |  | (1.423) |  |
| Native American | -0.174 | 0.182 | 2.091 | 0.449 |
|  | (0.130) |  | (2.763) |  |
| White | 0.045 | 0.326 | 1.110 | 0.256 |
|  | (0.046) |  | (0.977) |  |
| Other Race | -0.041 | 0.672 | 4.159 | 0.053 |
|  | (0.097) |  | (2.146) |  |
| Constant | -1.131\*\*\* | 0.000 | 11.814\*\*\* | 0.000 |
|  | (0.082) |  | (1.758) |  |
| *R*2 | 0.31 |  | 0.13 |  |
| *N* | 2317 |  | 1973 |  |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Figure S2, below, shows the modeled effect of hostile sexism on feeling thermometer ratings of each individual candidate, separately for both (5-item) measures of hostile sexism (Study 2).

**Figure S2. Predictive Validity for Individual Politicians**



**Table S6. Analysis of Perceptions of Sexual Misconduct Allegations (Study 2)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Credibility |  | Credibility |  |
|  | Beta (SE) | *p* | Beta (SE) | *p* |
| Hostile Sexism (5-item) | -0.732\*\*\* | 0.000 |  |  |
|  | (0.130) |  |  |  |
| Hostile Sexism (11-item) |  |  | -1.188\*\*\* | 0.000 |
|  |  |  | (0.173) |  |
| Revised Scale | -0.128 | 0.350 | -0.195 | 0.172 |
|  | (0.137) |  | (0.143) |  |
| Revised x Hostile Sexism (5) | 0.262 | 0.298 |  |  |
|  | (0.252) |  |  |  |
| Revised x Hostile Sexism (11) |  |  | 0.400 | 0.149 |
|  |  |  | (0.277) |  |
| Partisan Identity | -0.037\* | 0.012 | -0.028 | 0.058 |
|  | (0.015) |  | (0.015) |  |
| Ideology | -0.027 | 0.137 | -0.020 | 0.280 |
|  | (0.018) |  | (0.019) |  |
| Political Interest | 0.147\*\*\* | 0.000 | 0.143\*\*\* | 0.000 |
|  | (0.024) |  | (0.025) |  |
| Education | 0.039\* | 0.039 | 0.038\* | 0.041 |
|  | (0.019) |  | (0.019) |  |
| Age | 0.009\*\*\* | 0.000 | 0.009\*\*\* | 0.000 |
|  | (0.002) |  | (0.002) |  |
| Male | -0.060 | 0.283 | -0.045 | 0.418 |
|  | (0.055) |  | (0.056) |  |
| Asian | 0.362\*\* | 0.009 | 0.367\*\* | 0.007 |
|  | (0.138) |  | (0.137) |  |
| Hispanic | 0.354\*\* | 0.006 | 0.371\*\* | 0.004 |
|  | (0.128) |  | (0.128) |  |
| Native American | 0.628\* | 0.017 | 0.622\* | 0.018 |
|  | (0.263) |  | (0.262) |  |
| White | 0.288\*\* | 0.002 | 0.308\*\* | 0.001 |
|  | (0.094) |  | (0.094) |  |
| Other Race | 0.633\*\* | 0.003 | 0.616\*\* | 0.004 |
|  | (0.210) |  | (0.212) |  |
| Constant | 3.031\*\*\* | 0.000 | 3.172\*\*\* | 0.000 |
|  | (0.166) |  | (0.172) |  |
| *R*2 | 0.11 |  | 0.12 |  |
| *N* | 1554 |  | 1531 |  |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Question Wording of the Benevolent and Modern Sexism Scales**

Scale versions below are those tested in Schaffner (2021). The Benevolent Sexism scale matches the original items proposed by Glick and Fiske (1996) with some small wording changes. The items in the Modern Sexism scale differ a bit in wording from the original items proposed by Swim et al. (1995) but are used in the Voter Study Group’s VOTER survey and are similar to items used in the ANES.

Benevolent Sexism

A good woman should be set on a pedestal by her man

Women should be cherished and protected by men

Men should be willing to sacrifice for the women in their lives

In a disaster, women ought not necessarily be rescued before men

Women, compared to men, tend to have a superior moral sensibility

Many women have a quality of purity that few men possess

Women tend to have a more refined sense of culture and good taste

Every man ought to have a woman whom he adores

Men are complete without women

A man is not truly complete unless he has the love of a woman

People are often truly happy without a member of the other sex

Modern Sexism[[3]](#footnote-3)

Women often miss out on good jobs because of discrimination

Women who complain about harassment cause more problems than they solve

Sexual harassment against women in the workplace is no longer a problem

Increased opportunities for women have improved quality of life

1. \*Corresponding author. [↑](#footnote-ref-1)
2. In Figure S2, we show results for each candidate separately. There is no evidence that any individual candidate is driving the results. [↑](#footnote-ref-2)
3. Conceptually, Glick and Fiske (1996, 509) – the creators of the hostile sexism scale – note modern sexism is a better fit for studying “general political stances” and “gender-related political attitudes” (e.g., recognizing societal-level bias against women, public policy stances, or views about feminists) than hostile sexism and the ASI, which they argue is more applicable to “interpersonal relationships between men and women.” Yet, Schaffner (2021) finds the two are similar. [↑](#footnote-ref-3)