"Gender Stereotypes" 25 Years Later: A Replication of Huddy and Terkildsen

ANONYMISED AUTHOR(S)   *Anonymised Institution(s)*

*H*u*ddy and Terkildsen's seminal 1993 paper "Gender Stereotypes and the Perception of Male and Female Candidates" provided evidence that voters expect female candidates to be more competent on "compassion" issues and men more competent on military and defense issues. Huddy and Terkildsen find that voters rely on gender-trait stereotypes – gender-based personality traits – more than they do on gender-belief stereotypes – differences based on candidates' biological gender. We successfully replicate the main findings of the Huddy and Terkildsen paper using a set of participants in 2017. We find only small differences between our participants and the original study. In particular, 2017 participants seem to be even more likely to rely on trait stereotypes, indicating that current college students may be even more willing to separate biological gender from traits when making assessments of political candidates.*

Word Count:
3911 (errors:4)

INTRODUCTION

Twenty-five years ago, Huddy and Terkildsen (1993) found evidence that voters relied on gender stereotypes when assessing candidates for public office. Innovative at its time, the experiment was formative to a long literature on gender and candidate traits. Indeed, Google Scholar indicates the paper has been cited more than 900 times since its publication, including more than 25 times in the first three months of 2018. Since that time, however, much has changed. The share of women Senators has risen from 6% in 1993 to 21% in 2017.Furthermore, Hillary Clinton's historic presidential nomination has increased the political interest and ambition (Bonneau and Kanthak (2018)) of the women coming after her. It is time, then, to revisit the Huddy and Terkildsen (1993) results. The current project represents an attempt to replicate, as closely as possible, the Huddy and

Anonymised submission.

This is a manuscript submitted for review.

Terkildsen study on a set of participants in 2017.

In their study, Huddy and Terkildsen (1993) exposed college-aged subjects to a series of vignettes in which researchers manipulated the candidates' sex (by using either a male or a female name) and the candidates' traits (by describing them as having either feminine or masculine qualities), finding that subjects in the early 1990s saw candidates described as having more feminine traits as being better on "compassion" issues, whereas they saw candidates with more masculine traits as better able to handle the economy and the military. Further, female candidates were perceived as more competent on compassion issues, but there were no effects on other issues. Our replication, then, exposed current college students to the same vignettes to determine how the results may have changed over the ensuing 25 years.

Remarkably, many of their core findings have stood the test of time: Voters continue to associate traditionally masculine traits with better economic management skills, and still infer that female candidates' beliefs are more liberal, Democratic, and feminist. Equally interesting are the findings which have changed: Masculine-linked traits are no longer associated with greater competency in military matters, and male candidates are perceived to be less competent in economic issues after controlling for the effect of candidate instrumentality. In sum, our replication produces a compelling sketch of the nuanced nature of temporal shifts in gender perceptions, which can be best described as sticky but shapeable.

OUR REPLICATION

Following Huddy and Terkildsen (hereafter, H&T), we used experimental survey data of students evaluating a hypothetical political candidate. Subjects were first shown a candidate vignette, taken almost verbatim from the original article, describing the candidate's profession, age, state of residency, their longtime activism in politics, and offering a number of traits that colleagues have used to describe the candidate.[1] Students were then tasked with inferring positions of the candidate, rating

---

[1]In the original study, subjects were told that the candidates were from the neighboring state of Connecticut; in our study, we used the neighboring state of Ohio, in an attempt to replicate the "neighboring state" aspect of the experiment, rather than blindly staying with Connecticut. Otherwise, the vignettes were identical.

their perceptions of the candidate's competency on a number of issues, and the degree to which they associated the candidate with a variety of gendered personality traits. The positions, issues, and traits were all taken directly from the original study and were not updated, consistent with our desire to test whether the exact same gender traits, roles and competencies were still applicable. After completing these sections, students answered a set of demographic questions before concluding. Data was collected on from March 21 to April 18, 2017 using Qualtrics.

A total of three hundred and eighty-two undergraduate students at the University of Pittsburgh participated in the study in order to fulfill course requirements. The average age of our participants was 20, though subjects ranged in age from 18 to 41. 45% of our sample was in their first year of college, 28% in their second year, 17% in their third year and the remaining 8% were in their fourth year. 49% of students were political science majors, 5% were psychology majors, 8% were undeclared and the final 38% listed other for their major. 80% of the students identified as white, 3% Hispanic, 3% African-American, and 10% Asian. The sample was evenly split between men and women. The sample skewed Democratic, with almost 60% identifying as a Democrat versus only 20% identifying as Republicans.

Consistent with the original experiment, we utilized a 2x2x2 design. Subjects were randomly assigned to one of eight treatments, manipulating the candidate's gender (male or female), gender-linked traits (masculine or feminine) and the level of office being sought (national or local). As described above, we employed the same candidate names, candidate description, and traits as the original experiment.

Like H&T, the results from the main model demonstrate greater support for the gender-linked personality traits hypothesis, and we find only limited support for the belief stereotypes hypothesis. Overall findings notwithstanding, there are some minor differences in the final model, as well as some clear deviations from H&T's findings in other sections of the results. We follow the structure of H&T, noting when our results differ from the earlier work.

Trait Stereotypes

Like H&T, we begin by considering whether or not treatment successfully manipulated the participants' perceptions of the candidates. Regardless of whether each participant was treated with a male or female

candidate, we expect the feminine or masculine traits to imbue the candidates with gendered traits beyond those in the description. To test this, we create two scales[2] of inferred traits that correspond with the manipulated traits in the treatment. The first consists of seven typically feminine traits (warm, emotional, gentle, talkative, feminine, cautious, sensitive). We ask participants "How well do each of the following describe [the candidate]?" These seven items compose the *warmth/expressive* scale (α = .82). To test for masculine traits, we construct the *instrumentality* scale, which includes nine traits: assertive, stern, self-confident, coarse, masculine, tough, active, aggressive, rational (α = .75)[3]. We can be confident, then, that like in the original, we were able to overturn usual biological gender-based stereotypes.

As in the original, we employed analysis of variance (ANOVA) models to test the effects of the gender and trait treatments on average values of the inferred traits scales. For the *warmth/expressiveness* scale, the traits treatment was successful in manipulating participants' inferred traits for candidates (F[1, 361] = 59.30; p < .01).[4] The mean score for *warmth/expressiveness* was 8.11[5] for those treated with masculine candidates and 11.00[6] for those treated with feminine candidates (regardless of candidate gender). Meanwhile the gender of the candidate did not have a statistically significant effect on further inferred traits (means: 9.3[7] for female candidates and 9.78[8]). This demonstrates that the treatment of candidate traits was effective enough to overcome the potential gender-based stereotypes.

Similarly, the traits treatment had a significant effect on the *instrumentality* scale (F[1, 361] = 46.04; p < .01) with respondents reporting an average score of 11.74[9] for feminine candidates and 13.78[10] for masculine ones, independently of candidate genders. In this case, there was also a significant, albeit less so, effect (F[1, 361] = 4.79; p < .05) from the candidates' genders on the *instrumentality* scale

---

[2]We opted to use the same words as H&T, which produced scales that were very similar in internal coherence. The scales are standardized to range from 0-20.

[3]α = .83 in H&T

[4]F[1,289] = 63.83; p < .01 in H&T

[5]9.69 in H&T

[6]12.13 in H&T

[7]10.51 in H&T

[8]11.27 in H&T

[9]10.68 in H&T

[10]13.34 in H&T

4

(mean = 12.44 versus 13.09). Thus, we were less successful in overturning this scale than the first. This occurred in H&T, as well, but resulted from the *warmth/ expressiveness* scale in their case.

The replication, then, demonstrates the effectiveness of the gender-stereotyped traits of candidates in overturning potential candidate-gender. It shows that respondents understood the personality traits as somewhat independent from gender, enabling us to confidently proceed with tests of how gender and traits affect assessments of candidates within different issue areas.

Issue competency. Like H&T, we test how biological gender and traits match up on perceived competency in four policy areas. H&T argue that candidate assessments based on gendered traits (regardless of gender) would place more faith in masculine candidates for competence with military issues and feminine candidates for competence with compassion issues, but they made no predictions regarding the economy and women's issues. *Military* competency is a single item that asks respondents how their candidate would handle a military or police crisis. *Compassion* issues are on a scale of items dealing with children, poor people and the elderly (α = .81).[11] *Economic* issue competency includes three items related to budgets and finances (α = .69).[12] *Women's* issues consist of ability to manage gender inequality and abortion policy (α = .81).[13]

Like H&T, the ANOVA tests for differences in means among treatment groups provides mixed but overall positive support for the trait hypothesis. We also find that masculine traits in candidates (regardless of gender) result in greater perceptions of competency for military and police issues; however, our results more clearly promote the trait approach than that of H&T, as male candidates (regardless of traits) are also seen as more competent under H&T but not in our results.[14] As with H&T, both feminine candidates (regardless of gender) and female candidates (regardless of traits) are more seen as more competent with respect to compassion issues, which provides mixed support for the trait approach. Similarly, we also found that female candidates (but neither feminine nor masculine candidates) were seen as more competent regarding women's issues. While H&T predicted and found null results for both gender and traits on economic issue competence, we found that masculine

---

[11]α = .90 in H&T

[12]α = .69 in H&T

[13]α = .75 in H&T

[14]The significance criteria for these treatment groups come from ANOVA outputs. 1.

candidates are seen as more competent in economic issues. Finally, we corroborate H&T's finding that neither masculine nor feminine candidates are seen as more competent to deal with women's issues and that female candidates are seen as more competent in this area. Overall, both our results and H&T's results in the ANOVA models provide initial support for the claim that candidate traits, and not solely their genders, are important in explaining issue competency.

TABLE 1. Differences in Issue Competency & Political Beliefs by Candidate Gender and Traits

	Candidate Gender		Candidate Traits	
	Female	Male	Feminine	Masculine
Issue Competency:				
Military/Police(1-4)	3.016	3.006	2.894	3.127**
Economy(3-12)	9.361	9.220	8.989	9.591***
Compassion(4-16)	12.650	10.930***	12.610	11.000***
Women's issues(2-8)	6.383	4.949***	5.726	5.630
Political Beliefs:				
Democrat (1-3)	2.366	1.695***	2.006	2.066
Liberal (1-5)	3.355	2.661***	3.034	2.994
Support Feminism (1-4)	3.262	2.621***	3.000	2.895*
Observations	360		360	
Note:			*p<0.1; **p<0.05; ***p<0.01	

One distinction between our findings and those of H&T at this point in the analysis is the lack of significant interaction terms among the issue competency variables. For example, H&T finds significant differences between male and female treatment groups and between masculine and feminine treatment groups. Beyond this, the interaction term showed a significant increase in perceived competency among feminine men compared with masculine men in regards to compassion issues. The same significant increase in competency occurred for feminine, male candidates regarding women's issues (compared with masculine male candidates); however, for female candidates, it was being more *masculine* that benefited their perceived competency, exhibiting a diverging effect of masculinity between the male and female candidates. In contrast, our results demonstrate only significant interaction terms in the variables discussed in the following section (beliefs). In addition to being interesting findings, H&T's results demonstrate the confounding relationships that are present when the treatment groups are viewed in isolation of the inferred traits that respondents derive from the treatments. In a later section,

we shift the analysis from the treatment groups to the inferred traits (and beliefs) that resulted from respondents' understandings of the gender- and trait-based treatments.

Belief stereotypes

According to the belief approach outlined by H&T, female candidates (regardless of masculine or feminine traits) are presumed to hold beliefs that are more liberal, more Democratic and possibly more feminist than male candidates. Female candidates' presumed belief traits, in turn, provide them with the issue competencies that are associated with those beliefs. Under the *belief stereotypes* hypothesis, female candidates (through their beliefs) will have greater expertise in compassion and womens' issues, while male candidates will be associated with the opposite beliefs (more conservative, more Republican and less feminist) and will be perceived to be more competent than female candidates with issues related to the military/police and the economy.

Our findings in Table 1 are similar to those of H&T in that female candidates were seen as more being more liberal, more Democratic and more feminist than men. Similarly, female candidates were seen to have higher competency in both compassion and women's issues, lending strong support to the belief approach. Our results are less ambiguous, because, unlike in H&T, candidates' traits did not predict greater Democratic beliefs and feminist beliefs approach significance only as a result of candidates' traits. In addition to the main effects reported in Table 1, our results include several significant interaction terms under the inferred political beliefs in the ANOVA models: While feminine and masculine females are rated more highly as supporters of feminism than all males, masculine males are rated significantly lower than feminine males regarding feminism, and these relationships are paralleled with feminine males being seen as more liberal than masculine males. Regarding partisan beliefs, feminine females are seen as significantly more Democratic than masculine females, even though all female candidates are seen as significantly more Democratic than all males.

As discussed earlier in the results section, there was an imperfect correspondence between the treatment of gendered traits (masculine/feminine) and inferred candidate traits (using the warmth/expressiveness and instrumentality scales), demonstrated by the fact that male candidates were also seen as more instrumental, even after controlling for masculine traits. H&T also faces this problem, and

TABLE 2. Regression Results

	(1) Military	(2) Economics	(3) Compassion	(4) Women's Issues	(5) Compassion	(6) Women's Issues
Gender(Male)	-0.241	-0.316**	-0.747***	-1.196***	-0.649***	-1.337***
	(0.188)	(0.144)	(0.172)	(0.208)	(0.176)	(0.202)
Instrumental	5.043***	4.021***	-0.0432	2.535***	-0.0639	2.662***
	(0.564)	(0.467)	(0.572)	(0.679)	(0.568)	(0.685)
Warmth/ Express.	0.177	0.125	3.947***	1.902***	3.903***	1.951***
	(0.440)	(0.361)	(0.438)	(0.499)	(0.434)	(0.517)
Democrat	-0.190	0.180	0.486	0.488		
	(0.420)	(0.340)	(0.362)	(0.444)		
Liberal	-0.197	-0.740	0.892	0.239		
	(0.752)	(0.608)	(0.610)	(0.812)		
Pro-feminist				2.970***		
				(0.750)		
Liberal+Dem. +Fem.					1.902***	2.745***
					(0.450)	(0.529)
Constant	4.598***	5.591***	5.022***	2.504***	4.595***	3.286***
	(0.529)	(0.421)	(0.510)	(0.701)	(0.505)	(0.588)
Observations	361	361	360	361	360	361
R-squared	0.189	0.197	0.316	0.343	0.328	0.327
Note:						Robust standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

the solution lies in changing the independent variable from the treatment groups to the inferred traits and beliefs that result from the treatment groups. We also performed a simple bivariate test of these relationships (not reported), which shows correlations between the four issue areas and the inferred characteristics (both traits and beliefs). Like H&T, we find positive correlations between the expected female candidate beliefs (Democratic, liberal and feminist) and expected female competency (compassion and women's issues). The bivariate relationships between male candidate beliefs (Republican, conservative and less feminist) did not correspond well to those issues (military and economic), which is also the case to a lesser degree in H&T. The inferred candidate traits (instrumentality and warmth) are strongly correlated with their expected issue areas (military and compassion). The following section offers a clearer test to differentiate between the effects of the inferred traits and beliefs and the issue competencies.

Traits versus Beliefs as Source of Issue Stereotypes

In each test so far, the results have demonstrated support for both the traits and belief hypotheses; however, there is overlap in the results. By shifting from treatment categories to inferred beliefs and traits, as well as by including all of the variables of interest in the same model for each issue area, we are better able to test the differential influence of the trait and belief approaches. H&T finds greater support for the trait approach, and our results in Table 2 corroborate those findings. Following H&T, all independent variables have been rescaled to range from zero to one, and all dependent variables range from zero to ten. Models 1-4 are those directly replicated from H&T. Models 5-6 combine the belief traits into a scale, so that the reliability of the inferred belief traits measurement is more similar to that of the inferred personality traits, which were created using seven- and nine-item scales. Because these three beliefs were correlated (average coefficient = 0.63), they form a scale that is as reliable as the inferred trait scales (α =.83).

It is first worth noting that the only significant coefficient from the inferred beliefs variables in Models 1-4 is inferred pro-feminist views effect on competency in women's issues. This coefficient is large, strongly significant (B=2.5, $p > .001$) and is one of the predictions under the belief hypothesis. H&T found some support for the belief hypothesis: Democratic and liberal beliefs' were positive and significant in the compassion issues model – neither of which approaches significance in our results from Models 1-4. However, Models 5 and 6 show that the belief traits become highly significant and positive when combined in a scale (Lib.+Dem.+Fem.). This provides some support for the beliefs approach, and poses the question of whether H&T's results might have suffered from comparing single-measure belief measures with reliable traits scales. The military and economic issues models were unchanged by including the belief scale, suggesting that, even with this finding, the belief approach has limited support, overall.

Our results corroborate the direction and significance of the coefficients for the inferred instrumentality scales on competence in military issues, and, like H&T, we have positive and significant coefficients on economic issues and women's issues, even though these relationships were hypothesized to be null. We corroborate H&T's hypothesized coefficients for the warmth scale on both compassion and women's issues, and there is no longer a significant negative relationship between instrumentality

and compassion issues.

In H&T's model, candidate gender was "...included in all analyses to determine the success of traits and beliefs in eradicating customary gender differences" (H&T 1993: 136). Table 2 shows that male candidate gender has a significant and negative effect on economic issue competence after controlling for the inferred traits, and it is insignificant on military issue competence. This marks an interesting divergence from H&T's results, in which male candidate-gender had a positive and significant effect on military issue competence, and a null effect on economic issues. Compared with H&T, our larger coefficients for instrumentality and the larger R^2 statistics in Models 1 and 2 suggest that respondents in our experiment rely on candidate traits to a greater degree in making decisions about competency. In this sense, then, our modern subject pool may be more willing to infer characteristics from traits than were the original H&T pool 25 years ago.

In Figure 1, we plot the interaction between biological gender and traits. Similar to Figure 1 in H&T, we find that biologically female candidates are seen as more competent on women's issues than are biologically male candidates. Similar to H&T, we also see that male candidates with male traits are perceived as less competent on women's issues than are male candidates with female traits. From there, though, our results differ from those of H&T. Specifically, the line for female candidates is essentially flat, representing no difference in perceived competence on women's issues as female candidates take on male traits, whereas H&T found an increase in female candidates' competence on women's issues when the female candidate is depicted as having male traits. Also of note, in every circumstance, our subjects perceived every candidate as being more competent on women's issues than did the subjects in the original H&T study.

Like H&T, we find that male candidates were perceived as less competent in both compassion and women's issues. In order to ensure that including candidate gender (but not candidate traits assignment) did not affect the results, we tested the models without candidate gender, as well as without gender while adding a dummy for candidate-traits treatment assignment group. Neither of these modifications affected the instrumentality or warmth coefficients; however, the variables for Democratic and liberal became significant in their effect on compassion issues. Overall, candidate gender remains an important predictor of competence, even after controlling for beliefs and traits.

FIGURE 1. Interaction between Candidate Gender and Traits on Women's Issues
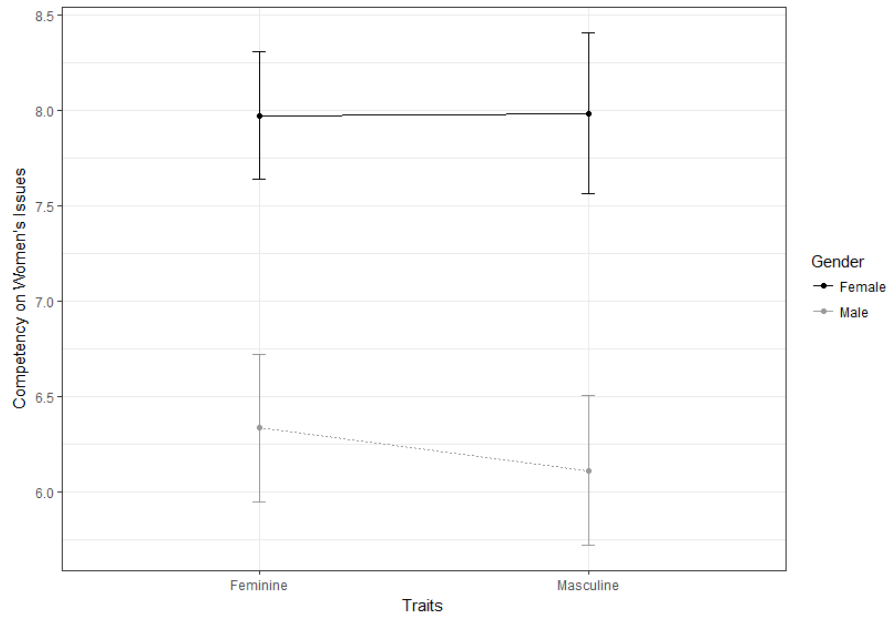
Note: 95% confidence intervals are shown with vertical lines.

DISCUSSION

As our results demonstrate, most of Huddy and Terkildsen's most important findings have stood the test of time. Most notably, voter use of gender-linked traits to assess candidate competencies is a remarkably persistent phenomenon. Equally compelling is the fact that the same gender-linked traits used in the original experiment yield similar effects today, suggesting that underlying assumptions about gender characteristics are static despite decades of social and cultural change. This result supports the conclusions generated in this line of research (e.g. Alexander and Andersen 1993; Prentice and Carranza 2002; Johns and Shephard 2007; Winter 2010; Dolan 2010). In contrast to some recent findings (Dolan 2014; Dolan and Lynch 2016), it appears that female voters continue to rely on gender cue to infer candidate beliefs, while voters of both genders use gender-linked traits to assess competencies.

In sum, gender still matters when voters are assessing political candidates, demonstrating the resiliency of stereotypes as a voter heuristic. In a modern political context characterized by substantially more female candidates than previous decades, this persistent effect is likely to have important consequences in the electoral landscape. Female candidates attempting to gain the support of liberal, Democratic feminists may find that their gender offers a significant advantage in their credibility on issues of concern to this group. Male candidates may find that they are less subject to the use of

stereotypes to infer beliefs about their positions, meaning they are more capable of retaining credibility when assuming liberal positions typically associated with female candidate. On the other hand, male candidates attempting to leverage their gender to obtain credibility with conservative voters will be unable to do so.

Performance of gender took center stage for both major party candidates for president in 2016. Performance of gender, from Bill Clinton proclaiming that he felt our pain to Hillary Clinton declaring that she would fight for us, will continue to play a role in elections. There is no reason to expect that it's role will diminish in the short term. Indeed, sex and gender will continue to act as a heuristic for voters. Our results indicate that gender operates among voters much as it did in 1993. Candidates, then, would be loathe to ignore the effect both biological gender and performance of gender have on voters' evaluations of their competence.

REFERENCES

Alexander, Deborah and Kristi Andersen (1993). Gender as a factor in the attribution of leadership traits. *Political Research Quarterly 46*(3), 527–545.

Bonneau, Chris W. and Kristin Kanthak (2018). Stronger together: political ambition and the presentation of women running for office. *Politics, Groups, and Identities 0*(0), 1–19.

Dolan, Kathleen (2010). The impact of gender stereotyped evaluations on support for women candidates. *Political Behavior 32*(1), 69–88.

Dolan, Kathleen (2014). Gender stereotypes, candidate evaluations, and voting for women candidates. *Political Research Quarterly 67*(1), 96–107.

Dolan, Kathleen and Timothy Lynch (2016). The impact of gender stereotypes on voting for women candidates by level and type of office. *Politics and Gender 12*(3), 573–595.

Huddy, Leonie and Nayda Terkildsen (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, 119–147.

Johns, Robert and Mark Shephard (2007). Gender, candidate image and electoral preference. *British Journal of Politics and International Relations 9*(3), 434–460.

Prentice, Deborah A and Erica Carranza (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of women quarterly 26*(4), 269–281.

Winter, Nicholas J. G. (2010). Masculine republicans and feminine democrats: Gender and americans' explicit and implicit images of the political parties. *Political Behavior 32*(4), 587–618.