



Kent Academic Repository

Chalmers, Jocelyn K. (2021) *Perceptions of Women in Authority Positions: The Role of Warmth and Competence*. Master of Science by Research (MScRes) thesis, University of Kent,.

Downloaded from

<https://kar.kent.ac.uk/88012/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.22024/UniKent/01.02.88012>

This document version

UNSPECIFIED

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Title: Perceptions of Women in Authority Positions: The Role of Warmth and Competence

Abstract: Warmth has been shown to disproportionately affect how we perceive women socially, resulting in different standards for men and women in authority roles that emphasize competence. I conducted 5 studies to determine whether warmth-related traits play a more central role in the evaluations of female lecturer and politicians than their male counterparts such that women are disproportionately “punished” in perceivers’ eyes for lacking warmth, but not rewarded for possessing it. In Studies 1 and 2 participants assessed the warmth, competence and overall quality of university lecturers and American politicians respectively; in Study 1 a lack of warmth was more integral to how women were evaluated than it was to how men were evaluated, but this was not replicated in Study 2. In Study 3 I analysed the content of Rate My Professor evaluations, which provided some evidence that women receive greater scrutiny than men on warmth-related dimensions. Studies 4 and 5 were content analyses of American political media sources; while these studies did not provide evidence for my hypothesis, they did provide some evidence for an oppositional relationship between warmth and competence for women but not men. Overall, these studies provide mixed support for the idea that women in authority are viewed disproportionately negatively for lacking warmth, and set the basis for future research into the role of warmth in evaluating high-status women.

Number of pages: 63

Year of submission: 2020

Academic school: Psychology

Perceptions of Women in Authority Positions: The Role of Warmth and Competence

Jocelyn K. Chalmers

University of Kent

Contents

Acknowledgments	3
Abstract.....	4
Introduction.....	5
Study 1	14
Method.....	15
Results	17
Discussion.....	20
Study 2	21
Method.....	22
Results	24
Discussion.....	27
Study 3	29
Method.....	30
Results	32
Discussion.....	34
Study 4	36
Method.....	36
Results	38
Discussion.....	39
Study 5	40
Method.....	41
Results	42
Discussion.....	42
General Discussion	43
References.....	55

Acknowledgments

I would like to thank my supervisor, Professor Robbie Sutton, for all of the guidance he provided me in conducting this project. I would also like to acknowledge the other members of the University of Kent's School of Psychology who generously provided me with support and feedback. Finally, I would like to thank my friends and family for their immense support during the completion of this project.

Abstract

Warmth has been shown to disproportionately affect how we perceive women socially, resulting in different standards for men and women in authority roles that emphasize competence. I conducted 5 studies to determine whether warmth-related traits play a more central role in the evaluations of female lecturer and politicians than their male counterparts such that women are disproportionately “punished” in perceivers’ eyes for lacking warmth, but not rewarded for possessing it. In Studies 1 and 2 participants assessed the warmth, competence and overall quality of university lecturers and American politicians respectively; in Study 1 a lack of warmth was more integral to how women were evaluated than it was to how men were evaluated, but this was not replicated in Study 2. In Study 3 I analysed the content of Rate My Professor evaluations, which provided some evidence that women receive greater scrutiny than men on warmth-related dimensions. Studies 4 and 5 were content analyses of American political media sources; while these studies did not provide evidence for my hypothesis, they did provide some evidence for an oppositional relationship between warmth and competence for women but not men. Overall, these studies provide mixed support for the idea that women in authority are viewed disproportionately negatively for lacking warmth, and set the basis for future research into the role of warmth in evaluating high-status women.

Keywords: women, gender roles, authority, politics, academia, warmth, competence

Women are underrepresented in top positions within both politics (Manning & Brudnick, 2020) and academia (Mengel, Sauermann, & Zölitz, 2018), and tend to receive enhanced negative scrutiny within these fields. Previous research (e.g., Heilman & Okimoto, 2007; Sprague & Massoni, 2005) suggests this may be related to how men and women are differentially perceived with regards to the two traits most central to person perception: warmth (i.e., friendliness; trustworthiness) and competence (i.e., agency; efficiency). These dimensions are gendered in that masculine stereotypes align more closely with competence traits and feminine stereotypes with warmth traits. Competence is generally desired in figures of authority, resulting in unique challenges when women fulfill leadership roles, since they are expected to embody the (masculine) agentic expectations of a leader while still embodying the warm, communal expectations their gender role dictates (Rudman, Moss-Racusin, Phelan, & Nauts, 2012). Teaching is one authority role in which this set of challenges arises, and research shows that women are consistently rated more harshly than their male colleagues in student evaluations (Mengel, Sauermann, & Zölitz, 2018). Similarly, there is evidence that when women are elected to political office they experience harsher and more intense scrutiny than their male colleagues (Żemojtel-Piotrowska, Marganski, Baran, & Piotrowski, 2016). With this in mind, these studies aimed to determine whether warmth-related traits are more central to how women in academia and politics are evaluated compared to men, and specifically whether women are punished more readily for low levels of warmth and rewarded less readily for high levels.

Warmth and Competence

Research in social cognition has established two major dimensions which are universally linked to person perception: warmth and competence. The dimension of warmth encompasses traits such as helpfulness, friendliness, and trustworthiness, which are related to one's perceived

intent. Conversely, the dimension of competence is related to one's perceived ability, and involves traits such as intelligence, skill, and efficiency. Traits related to warmth aid in relationships with other people, while competence mainly impacts one's own success or failure (Fiske, Cuddy, & Glick, 2007).

Sometimes referred to as agency and communion, warmth and competence were initially thought of as two opposing poles of the same dimension. In recent research, there is debate over whether warmth and competence are orthogonal or whether perceptions of one influence perceptions of the other. Imhoff and Koch (2017) proposed a curvilinear relationship between agency and communion such that individuals very high or very low on agency are perceived as lacking in communion, while those with average levels of agency are perceived as the most communal. However, others have proposed that warmth and competence tend to be conceived of as directly oppositional, such that high levels of agency imply a deficit of communion, and vice versa (Heilman & Okimoto, 2007). Kervyn, Bergsieker, and Fiske (2012) resolve this discrepancy somewhat by suggesting that societal ingroups (e.g., men) are often seen as both warm and competent, while societal outgroups (e.g., women) are generally characterized as high on one dimension but low on the other, which they describe as a "hydraulic" relationship (p. 78).

These traits are related to gender in the sense that women's traditional gender roles emphasize warmth while men's traditional gender roles emphasize competence (Heilman & Okimoto, 2007). Gender roles are collections of attitudes, traits and behaviours that are more associated with one gender than the other, and that form the basis of society's expectations for men and women: the traditional conceptualization of the male gender role is related to agency, dominance and assertiveness, while the female gender role is associated with sensitivity, emotional expressiveness and communality (Levesque, 2011). These descriptive stereotypes (i.e.,

how women *are*) often align with prescriptive stereotypes (i.e., how women *should be*), the violation of which can lead to negative backlash. Due to the perceived complementary nature of men and women within society, defying one's own gender stereotype generally involves straying into the opposite gender's stereotype, such that behaviours valorized in men are punished in women, and vice versa (Heilman, 2001).

Prentice and Carranza (2002) take these ideas a step further by suggesting that individuals are not automatically perceived negatively for demonstrating opposite gender traits, but rather that they are severely socially punished for demonstrating *negative* traits associated with the opposite gender. For example, a woman can be intelligent and efficient while still being viewed positively, but if she displays arrogance – a negative, masculine quality – this is seen as far less desirable than a man displaying arrogance, or a woman displaying a negative feminine trait such as weakness. Practically, however, this is of little consolation given that a woman acting assertively will almost inevitably veer into what others perceive as arrogance on some occasions. Within the framework suggested by these authors, women who seem arrogant, domineering, or insensitive – traits very much at odds with the communal feminine gender role – pay a higher evaluative price than men.

While warmth has often been thought of as interchangeable with moral character, some recent research (e.g., Goodwin, Piazza, & Rozin, 2014; Brambilla, Rusconi, Sacchi, & Cherubini, 2011) has suggested that information based on morality (e.g., honesty) is processed distinctly from information based on the sociability-related aspects of warmth (e.g., friendliness). However, while traits related to sociability are clearly aligned with female gender stereotypes (Rudman & Glick, 1999), definitive research is lacking on whether morality as an isolated

dimension is relatively more favoured in one gender or the other, and further investigation is needed in this area.

Gender Roles and Authority

The agentic male gender role is much more congruent with what is expected of a person in authority than the female one; as a result, women in high-status domains such as politics and academia face a double bind as they try to overcome the lack of fit between their feminine gender role and the qualities desired in an authority figure (Rudman et al., 2012). This is addressed in role-congruity theory (Eagly & Karau, 2002), which suggests that men are automatically assumed to possess the necessary qualities to perform well in positions of leadership due to the high level of gender-role congruity, while women have to work harder to be viewed as qualified. Given that warmth and competence tend to be thought of as oppositional in women, if they try to overcome this lack of fit by enacting agency, this display of competence may come at the cost of their perceived warmth and provoke a backlash related to behaving in an insufficiently feminine manner.

Several studies have yielded support for this theory. Rudman and Glick (2001) found that agentic female job applicants were judged as less likable and hireable than equally agentic male candidates, suggesting the feminine prescription for communality may prevent women from attaining high-status positions. Heilman and Okimoto (2007) similarly found that successful female managers were disliked and interpersonally derogated more than men who were described identically, but added that offering information suggesting these women possessed communal qualities mitigated the negative reactions. This result seems to suggest that the perceived lack of desirable feminine qualities is central to the bias against agentic women. This strong prescription for warmth-related traits in women was also found by Prentice and Carranza

(2002) in their interrogation of traits both desired in and typical of Princeton undergraduates. These authors found that while female students were perceived to possess the competence-related traits necessary for any person to succeed at a top university, they were also expected to possess a litany of warmth-related traits which were irrelevant to their academic success and not expected of men.

Perhaps due to the perceived role incongruity these traits contribute to, women continue to be underrepresented in both academia and politics. Men and women enter graduate school in equal numbers, but women are underrepresented in academic careers, particularly in higher-up positions such as senior professorships (Mengel et al., 2018). When it comes to politics, a woman has still never been elected president of the United States as of the year 2020, and still only comprise approximately one quarter of US congress and the senate (Manning & Brudnick, 2020). When women do assume these roles, people frequently discount their contributions and are less willing to be influenced by them, particularly if they strongly violate gendered expectations (Sprague & Massoni, 2005). Women behaving in an agentic manner can elicit discomfort and lead to negative characterizations such as “bitchy,” “selfish,” and “ice-queen.” A poignant example is Senator Hillary Rodham Clinton, who was consistently portrayed as untrustworthy, cold and corrupt in her 2016 run for the United States Presidency (Bligh, Schlehofer, Casad, & Gaffney, 2012).

Teaching Evaluations

“We cannot set aside the social relationships of the larger world— a world in which classifications of gender, race, and class are among the most paramount—as we take up the more temporary relationship of professor and student.” (Rakow, 1991, p. 10)

Research on women in academia has consistently shown that women receive poorer teaching evaluations from their students than men do, despite no evidence they are less effective teachers. For example, Mengel and colleagues (2018) examined several thousand evaluations of male and female university faculty by their students and found that female faculty were rated more negatively than their male peers even though students' grades and self-reported study hours were unaffected by the gender of their instructor. MacNell, Driscoll and Hunt (2015) took this a step further and found direct evidence of gender bias in evaluations of instructors in an online course. These authors had instructors each pose as two different gender identities to control for the influence of other factors; students rated the male identity significantly higher than the female identity, regardless of the actual gender of the instructor. These authors also found that lecturers perceived as male were rated higher on all interpersonal measures (e.g., measures of respect, enthusiasm, and warmth), even though both instructors employed the same level of interpersonal interaction in the online course. The authors conjectured that female instructors were punished for the absence of these interpersonal traits but not rewarded for their presence, while male instructors were rewarded for going "above and beyond".

Crucially, the dimension of warmth appears more related to the evaluation of female (compared to male) lecturers. El-Alayli, Hansen-Brown, and Ceynar (2018) used the term "academic momism" (p.137) to describe how female lecturers were expected to take on interpersonal tasks like helping students cope with their stresses and insecurities, as well as providing feedback in a gentle manner to avoid being seen as too harsh. These authors suggested this was the result of women being perceived as more communal and therefore being held to a higher standard with regards to nurturing, warmth-related behaviours. Sprague and Massoni (2005) investigated the role different traits play in bias towards female faculty by conducting a

qualitative analysis of the words that college students used to describe their teachers. Words that meant *Compassionate*, *Sensitive*, and *Giving* – which imply a large degree of emotional labour – were only used to describe women teachers, while men were described with positive attributes that did not require the same amount of time and investment (e.g., funny, spontaneous). Students reported getting more personal time and attention from women teachers than from men, and yet were more likely to say women instructors were not available enough, indicating an expectation that women attend to their individual needs. When women failed to display the appropriate amount of warmth, they were described with words such as bitch, bitchy, bitch toward male students, witch, and feminazi. The authors proposed that these words indicate an attack on women not just as teachers, but rather as people; there were no similarly insulting and gender-specific terms used to describe men teachers. Mitchell and Martin (2018) also uncovered an emphasis on different criteria for male and female lecturers through an analysis of student evaluations. Not only were female lecturers evaluated more poorly than their male counterparts in identical courses, but women’s personality and appearance were referenced to a greater degree and women were referred to as “teacher” rather than “professor” more frequently than men, suggesting both an emphasis on women’s warmth and a tendency to undervalue their competence.

Female Politicians

Politics is another area of prestige and authority in which bias against women is consistently reported. Although men and women win elections at equal rates in the United States, there are far fewer women in office, and female candidates tend to have more political experience, stronger professional profiles and higher levels of education than the men they run against (Fulton, 2012). As of 2020, a woman has never been elected US president, and women

only comprise 23.8% of Congress and 26.0% of the Senate despite historic recent gains (Manning & Brudnick, 2020).

In one study investigating gender bias in American politics, Paul & Smith (2008) assessed perceptions of specific politicians from the United States. Despite these authors' attempts to systematically match the male and female politicians on their qualifications, participants consistently rated the two women – Senator Hillary Clinton and Senator Elizabeth Doyle – as significantly less qualified than their male peers (former senator John Edwards, New York City Mayor Rudy Giuliani, and Senator John McCain). Bligh and colleagues (2012) chose to assess the bias against female politicians through another important lens: political media coverage. These authors found that the framing of a news story has a particularly strong impact on the public's perception of the warmth and likability of the woman in question. When an article focused on a female politician's personality, she was scrutinized for evidence of warmth. If the article was positively framed, she was judged to be warm and aligned with female stereotypes (i.e., more warm than competent), while a negatively framed article resulted in her being regarded as cold and unlikable despite no direct mention of her personality traits. Unfortunately, these authors did not include male politicians in their assessment, rendering it difficult to conclude whether or not this effect was related to the gender of the politician.

Consistent with the notion that men and women are not punished to the same degree for their transgressions, bias against female politicians has also been found when politicians break the rules. Żemojtel-Piotrowska and colleagues (2016) conducted a study on perceptions of politicians involved in scandals in which participants were presented with fictional male and female politicians said to be involved in situations involving corruption or sexual affairs. These authors found that women were assessed less favourably on dimensions of electability, morality

and competence than men who took part in the exact same unethical behavior. These researchers explained that the discrimination shown against female politicians was not a simple, linear effect: women were perceived just as favourably as men when they had not violated any expectations, but as soon as they were involved in immoral behavior they were judged far more severely than their male counterparts.

The Present Research

The literature I have reviewed indicates that women in leadership positions such as in academia and politics are held to different standards than men, particularly when it comes to traits related to warmth, which disproportionately affect how we perceive women socially. When women enact the role of a lecturer or politician, they are expected to balance warmth and agency in a way not demanded of men. With this in mind, these studies aimed to quantitatively determine whether warmth-related traits play a more central role in the evaluations of female lecturers and politicians such that women have to display greater levels of warmth than their male counterparts in order to be viewed in an equally positive manner.

In Study 1, I assessed university students' ratings of their male and female lecturers' warmth and competence traits and compared them with their overall evaluations of those lecturers. In Study 2, I used the same methods to assess US Democratic party supporters' perceptions of the male and female candidates running for the Democratic party presidential nomination. In Study 3, I used an existing database cataloguing word use in Rate My Professor reviews to compare the frequency of warmth- and competence-related words in reviews of male and female lecturers. In Study 4, I conducted a content analysis of articles from *The New York Times* and *The Washington Post* and compared the use of warmth- and competence-related words when describing male and female candidates for the Democratic presidential nomination;

in Study 5, I conducted the same content analysis of articles from *The Wall Street Journal*, *The Boston Globe* and *CNN*.

In general, I hypothesized that warmth-related traits would be more central to the evaluations of female lecturers and politicians than those of their male counterparts, such that women would be rated more unfavorably than men when they lacked warmth, and warmth-related words would be used more frequently to describe women than men – particularly warmth-related words with a negative valence.

Study 1

This study examined students' evaluations of their university lecturers. Participants indicated their overall satisfaction with male and female lecturers, then rated them on perceived warmth along with other traits. I tested the hypothesis that there is a positive correlation between the perceived warmth of female instructors and their ratings by students such that women who are deemed warm are rated significantly more positively than their female colleagues who are not deemed to be warm. Specifically, I expected that as the rated warmth of female lecturers increased, it would relate less strongly to overall ratings. This reflects a pattern where women are rated unfavorably when they are perceived to lack warmth, but not rated especially favourably when they are warmer than expected. In contrast, I expected the correlation between warmth and overall evaluation of male lecturers to strengthen as warmth increases, reflecting a pattern where men are not particularly derogated for lacking in perceived warmth, but are valorized when they are warmer than expected. These hypotheses were preregistered on the Open Science Framework (OSF) (Chalmers, 2019).

Method

Participants.

Participants were 414 undergraduate psychology students at the University of Kent in their first or second year of study; 355 were female, and the mean age was 19.30 years old. They were recruited online through the University of Kent's Research Participation Scheme (RPS).

Procedure.

Participants completed an online survey. They were randomly assigned to either a male or female lecturer condition, and then presented with a list of the psychology lecturers who taught undergraduate modules in their year. Participants were asked to select from the list all lecturers whose classes they had attended and whom they felt able to evaluate, and then were randomly presented with one of the selected lecturers whose gender corresponded with the condition they were assigned to and told they would be evaluating that lecturer. They were first asked to indicate their familiarity with this lecturer on a 5-point Likert scale where 1 = *Not at all familiar* and 5 = *Very familiar*, then were asked to indicate how satisfied they were overall with the lecturer, how much they liked the lecturer as a person, and how effective the individual was as a lecturer on similarly formatted 5-point Likert scales. Next, they were asked if they would consider nominating the lecturer for a teaching prize on a scale from 1 = *Definitely not* to 5 = *Definitely yes*. Participants then rated the lecturer on warmth and competence by indicating the degree to which they met the student's expectations on 20 different traits related to these dimensions. Lastly, participants completed measures of sexism, religiosity, and right-wing authoritarianism (RWA) in order to explore the potential impact of these variables. Sexism was naturally included in this study to adjust for its potentially confounding role in gender-related perceptions; previous research has also found that high levels of religiosity and RWA are

predictors of sexism (Burn & Busso, 2005; Sibley, Wilson, & Duckitt, 2007), so these variables were adjusted for as well.

Measures.

Warmth/Competence. To assess the perceived competence and warmth of each lecturer, the items from Abele and colleagues' (2016) AC-IN scale were used. This scale contains 20 items divided into four categories: assertiveness (e.g., self-confident; $\alpha = .76$); competence (e.g., intelligent; $\alpha = .92$); warmth (e.g., friendly; $\alpha = .92$); and morality (e.g., fair; $\alpha = .88$). I altered the response scale such that participants indicated how they would rate the lecturer on these items "compared with what [they] expect from a lecturer" on a 7-point Likert scale from 1 = *Much less than I expect* to 7 = *Much more than I expect* ($\alpha = .94$).

Exploratory Variables. To assess sexism, participants completed a shortened form of the Ambivalent Sexism Inventory (Glick & Fiske, 1996). This scale involves participants indicating their level of agreement on a six-point scale ranging from 0 = *Disagree strongly* to 5 = *Agree strongly* with 4 statements assessing hostile sexism (e.g., "Women seek to gain power by getting control over men"; $\alpha = .90$) and 4 statements assessing benevolent sexism (e.g., "Women should be cherished and protected by men"; $\alpha = .83$) ($\alpha = .90$). To assess religiosity, a four-item scale was used (e.g., "How often do you attend religious services?"; Sullivan, 2001). Responses were recorded on a five-point scale ranging from 1 = *Not at all* to 5 = *A great deal* ($\alpha = .93$). To assess right-wing authoritarianism, a shortened form of the Right-Wing Authoritarianism (RWA; Altemeyer, 1981) scale was used. Participants responded their level of agreement with eight statements (e.g., "What our country really needs, instead of more "civil rights" is a good stiff dose of law and order") on a scale from 1 = *Strongly disagree* to 6 = *Strongly agree* ($\alpha = .73$).

Results

My main hypothesis, that the warmth of female lecturers would be significantly correlated with their evaluations by students and that this correlation would be stronger at low levels of warmth, was supported. A mean split was performed to test the differences between correlations above and below the mean level of warmth ($M = 5.04$) in female lecturers (89 participants above the mean; 116 below) and male lecturers (103 participants above the mean; 101 below). While there was no evidence that male lecturer's warmth was more strongly correlated with ratings above the mean, there was a much weaker correlation between men's warmth and overall ratings than between women's warmth and overall ratings at low levels of warmth. When warmth was below the mean, there was a significant difference between its correlation with perceived teaching effectiveness in men, $r(101) = .02, p = .49$, versus women, $r(116) = .35, p < .001, z = -2.16, p = .03$. The same was true for warmth's correlation with willingness to nominate for a teaching prize, which was non-significant in men, $r(101) = .07, p = .47$, but significant in women, $r(116) = .43, p < .001; z = -2.80, p = .005$, and for an overall evaluation variable combining satisfaction, liking, effectiveness, and prize nomination, which was also non-significant in men, $r(101) = .02, p = .14$, but significant in women, $r(116) = .44, p < .001; z = -2.34, p = .02$. For male lecturers, there was no significant difference in the warmth/evaluation correlation above and below the mean level of warmth. For female lecturers, the correlation between warmth and willingness to nominate for a teaching prize was significantly stronger below the mean, $r(116) = .43, p < .001$, than above the mean, $r(89) = .17, p = .12, z = 2.02, p = .04$.

While there was no significant difference in overall warmth and evaluation correlations between men and women, there was a significant difference in overall liking and evaluation

correlations. There was a significant positive relationship between liking and the combined evaluation variable in men, $r(206) = .69, p < .001$, and women, $r(206) = .86, p < .001$, but these correlations were significantly different from one another, $z = -2.56, p = .01$. Male lecturers were also liked better than female lecturers overall, $t(410) = 2.28, p = .02$.

When looking at conglomerate trait and evaluation relationships, there was no evidence of an overall bias against female lecturers. However, consistent with my hypothesis, opposing curvilinear patterns emerged for men and women. There was a significant convex quadratic relationship between an overall trait variable combining warmth, morality and competence and the combined evaluation variable in men, $F(2,201) = 37.23, p < .001$ (see Figure 1). (Note: An examination of the Mahalanobis distance scores indicated one multivariate outlier, but removing this from the analysis did not alter the significance of this relationship).

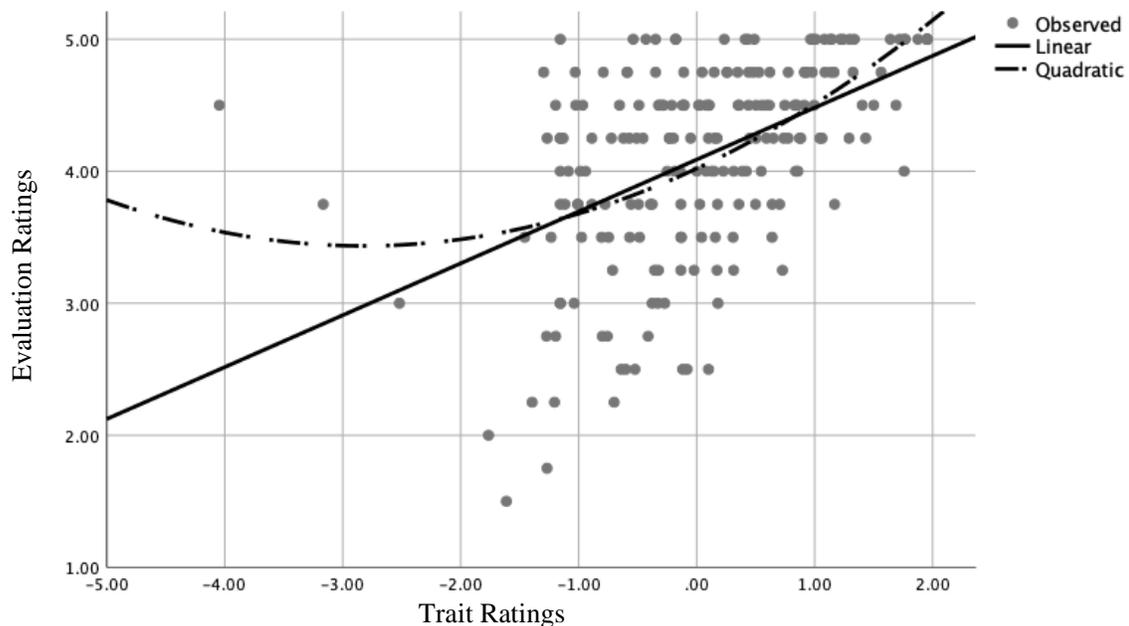


Figure 1. Overall trait and evaluation ratings in male lecturers.

In women, there was a significant concave quadratic relationship between the combined trait variable and the combined evaluation variable, $F(2,201) = 50.37, p < .001$ (see Figure 2).

(Note: An examination of the Mahalanobis distance scores indicated one multivariate outlier, but removing this from the analysis did not alter the significance of this relationship).

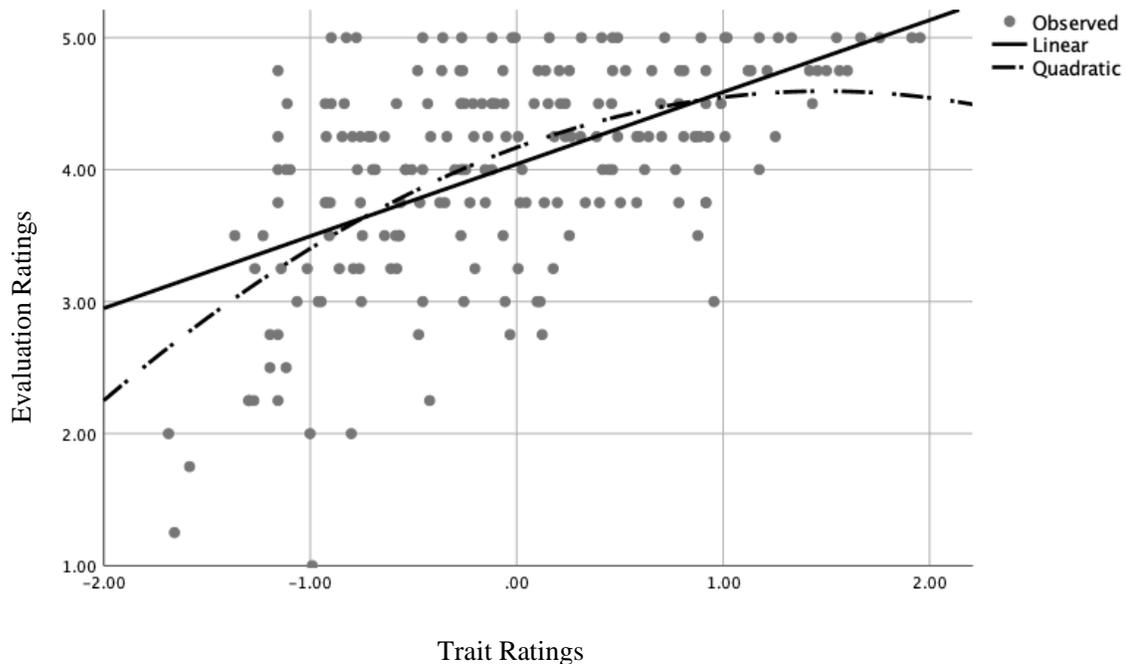


Figure 2. Overall trait and evaluation ratings in female lecturers.

There was somewhat of a ceiling effect with the combined evaluation variable in that scores clustered toward the high end of the measure. In the full dataset (male and female lecturers), 57 out of 412 participants who provided data for that measure indicated the highest possible value of 5 out of 5, which amounts to 13.8% of the sample indicating the maximum value the scale is set to measure. Comparatively, only, 1 out of 412 participants had a minimum score of 1 out of 5 on this variable, constituting 0.2% of the sample. This ceiling effect makes it more difficult to ascertain the true effect of the independent variables on this combined

dependent variable, because surveys like this create an artificially low ceiling such that the independent variable no longer has an effect on the independent variable (Garin, 2014).

There was no significant impact of any of the exploratory variables (sexism, religiosity, or right-wing authoritarianism) or of the gender of the participant on lecturer ratings.

Discussion

I hypothesized that the warmth of female lecturers would be significantly correlated with their evaluations by students and that this correlation would be stronger at low levels of warmth. My findings supported this: when lecturers were not deemed warm, the level of warmth of the female lecturers was much more central to how they were evaluated on several dimensions than it was for the male lecturers. Additionally, there was a significant difference in the warmth and evaluation correlation above and below the mean level of warmth for women, while for men there was not. Contrary to my hypothesis, however, there was no evidence that male lecturers' warmth was more strongly correlated with ratings when it was above the mean than female lecturers' was: the strongest correlations overall existed for women who were at low levels of warmth. While I cannot make causal arguments based on this finding, one interpretation is that women are punished more than men for failing to demonstrate warmth: when men are cold, it does not particularly impact how they are perceived, while for women it leads to negative perceptions. This supports previous research suggesting that warmth is more aligned with the feminine gender role, and thus is more central to how women are perceived.

Interestingly, while there was no significant difference in the overall warmth and evaluation correlations between men and women, there was a significant difference when it came to liking. A significant positive relationship between liking and evaluation existed for both men and women – as one would expect – but the correlation was significantly stronger for women,

and male lecturers were also liked better than female lecturers overall. Since this increased liking was not accompanied by increased ratings of warmth, competence or morality, one might conclude that the likeability of a male lecturer is simply not as dependent on the traits they embody as it is for a female lecturer. This is consistent with the idea that women may be violating the communal norms of their gender role when they take on a lecturer role, and as a result may have to work harder to be liked by students.

While there was no evidence of an overall bias against female lecturers when taking all traits and evaluations into account, consistent with my hypothesis, opposing curvilinear patterns did emerge for men and women. There was a convex quadratic relationship between traits and evaluations in men, and a concave quadratic relationship between the traits and evaluations in women. This suggests an overall trend where men's traits are not as correlated with overall evaluations below the mid-point, while women's traits are not as correlated with evaluations when they are above the mid-point. It is interesting to note that, due to the lack of overall differences in traits and evaluations, these trends would not have been revealed had I simply analyzed the results in a linear fashion. Using my novel non-linear approach, I were able to uncover the complex relationships that exist between traits and evaluations for men and women. In Study 2, I will attempt to replicate this effect in the political realm.

Study 2

This study attempted to replicate the results of the first study while examining perceptions of male and female candidates running in the United States Democratic Presidential Primary. Participants completed an overall assessment of male and female politicians, then rated them on perceived warmth along with other traits. Just as in Study 1, I tested the hypothesis that there is a positive correlation between the perceived warmth of female politicians and their

ratings such that women who are deemed warm are rated significantly more positively than their female colleagues who are not deemed to be warm. I did not expect this effect to occur with male politicians. Once again, these hypotheses were preregistered on OSF (Chalmers, 2020).

Method

Participants.

Participants were 287 Americans recruited through the online platform Prolific Academic; 162 were female, 123 were male, and 2 indicated their gender as Other. The mean age was 34.07 years old. Participants were pre-screened for a U.S. political affiliation of “Democrat” using Prolific Academic’s screening criteria.

Procedure.

Participants completed an online survey. They were randomly assigned to either a male or female politician condition, and asked to complete a series of demographic questions that included their level of political engagement and who they voted for in the 2016 general election. They were then presented with a list of 13 individuals formerly running for the Democratic Party’s presidential nomination: all 6 women (Elizabeth Warren, Amy Klobuchar, Kamala Harris, Kirsten Gillibrand, Marianne Williamson, and Tulsi Gabbard) and 7 men selected based on their all-time polling numbers (Bernie Sanders, Pete Buttigieg, Michael Bloomberg, Cory Booker, Beto O’Rourke, Julián Castro, Andrew Yang, and Tom Steyer). Presumptive Democratic nominee Joe Biden was not included due to concerns that his status as the “winner” of the contest might influence evaluations. Participants were asked to select from the list all candidates who they were familiar with and felt able to evaluate, and then were randomly presented with one of the selected politicians whose gender corresponded with the condition they were assigned to. They were first asked to indicate their familiarity with this politician on a 5-

point Likert scale where 1 = *Not at all familiar* and 5 = *Very familiar*, then were asked to indicate to what extent that person was a good representative of their interests, how qualified they were to be president, and how much they liked them as a person on similarly formatted 5-point Likert scales. They were also asked to assess the retrospective electoral chances of this politician in both the Democratic primary and the presidential election on a 5-point Likert scale where 1 = *Very low* and 5 = *Very high*, as well as their own willingness to vote for the politician in a primary from 1 = *Not at all willing* to 5 = *Very willing* both a) in a normal setting and b) if they did not have to take their electoral chances into account.

Participants then rated the politician on warmth, competence, morality and political skills by indicating the degree to which they possessed 16 different traits related to these dimensions. Participants also completed measures of sexism, religiosity, and right-wing authoritarianism in order to explore the potential impact of these variables. Lastly, participants were asked to indicate to what degree they thought women in politics were a) capable, b) likeable and c) electable compared to men on a 5-point Likert scale where 1 = *Much less than men* and 5 = *Much more than men*.

Measures.

Warmth/Competence/Morality. To assess the perceived competence, warmth and morality of each politician, I used items from a previous study by Brambilla and colleagues (2011). These authors created a scale using three morality traits (sincere, honest, trustworthy; $\alpha = .94$), three warmth traits (friendly, warm, likeable; $\alpha = .92$), and three competence traits (intelligent, competent, skillful; $\alpha = .92$). Participants indicated how they would rate the politician on these traits on a 5-point Likert scale where 1 = *Not at all* and 5 = *Very*. I chose to include morality

traits in this study based on Goodwin and colleagues' (2014) research suggesting moral character and social warmth traits are separable ($\alpha = .95$).

Political Skills. To assess skills related to success in politics, I used items from research by Fox and Lawless (2011). These authors assessed "Perceptions of Political Skills" using 6 attributes: knowledge about public policy issues; professional experience relevant to politics; public speaking skills; connections within the political system; fundraising ability; and self-promotion ability. I added "electability" to this list due to the frequency with which this attribute was mentioned in the run-up to the 2020 Democratic primary by both politicians and members of the media. Participants indicated to what extent the politician possessed these attributes on a 5-point Likert scale where 1 = *Not at all* and 5 = *A great deal* ($\alpha = .89$).

Exploratory Variables. The same measures were used to assess sexism (benevolent: $\alpha = .85$; hostile: $\alpha = .90$; total: $\alpha = .90$), religiosity ($\alpha = .94$), and right-wing authoritarianism ($\alpha = .81$) as in Study 1.

Results

My main hypothesis, that the warmth of female politicians would be significantly correlated with their evaluations and that this correlation would be stronger at low levels of warmth, was not supported. Once again, a mean split was performed to test the differences between correlations above and below the mean level of warmth ($M = 3.75$) in female politicians (77 participants above the mean; 68 below) and male politicians (82 participants above the mean; 60 below). Contrary to my previous results, there was no significant difference between men and women's warmth and overall evaluation correlations at either low or high levels of warmth. When warmth was below the mean, its only correlation that was significantly different between men and women was with representativeness. Surprisingly, warmth was significantly more

correlated with representativeness in men, $r(60) = .66, p < .001$, than in women, $r(68) = .34, p = .004, z = 2.41, p = .02$. The only correlation that was stronger for women at low levels of warmth was between liking and electability in the general election: this correlation was non-significant for men, $r(59) = -.024, p = .854$, but significant for women: $r(68) = .41, p = .001, z = 2.49, p = .01$. In the overall population, warmth was significantly more correlated with the overall evaluation variable (a combination of representativeness, liking, qualifications, electability, and willingness to vote for) in men, $r(142) = .80, p = .001$, than women, $r(145) = .66, p < .001, z = 2.46, p = .01$.

When looking at differences between means rather than correlations, the two significant differences were in the domains of qualifications and electability in the primary. Women were rated as significantly more qualified to be president than men, $t(285) = 2.39, p = .02$, and also as having significantly greater electoral chances in the democratic primary, $t(284) = 2.80, p = .005$.

When assessing conglomerate trait and evaluation relationships, there was no evidence of an overall bias against either male or female politicians. Unlike in Study 1, there were no opposing curvilinear patterns for men and women. There was a significant convex quadratic relationship between an overall trait variable combining warmth, morality, competence and political skills and the combined evaluation variable in men, $F(2,139) = 178.53, p < .001$ (see Figure 3).

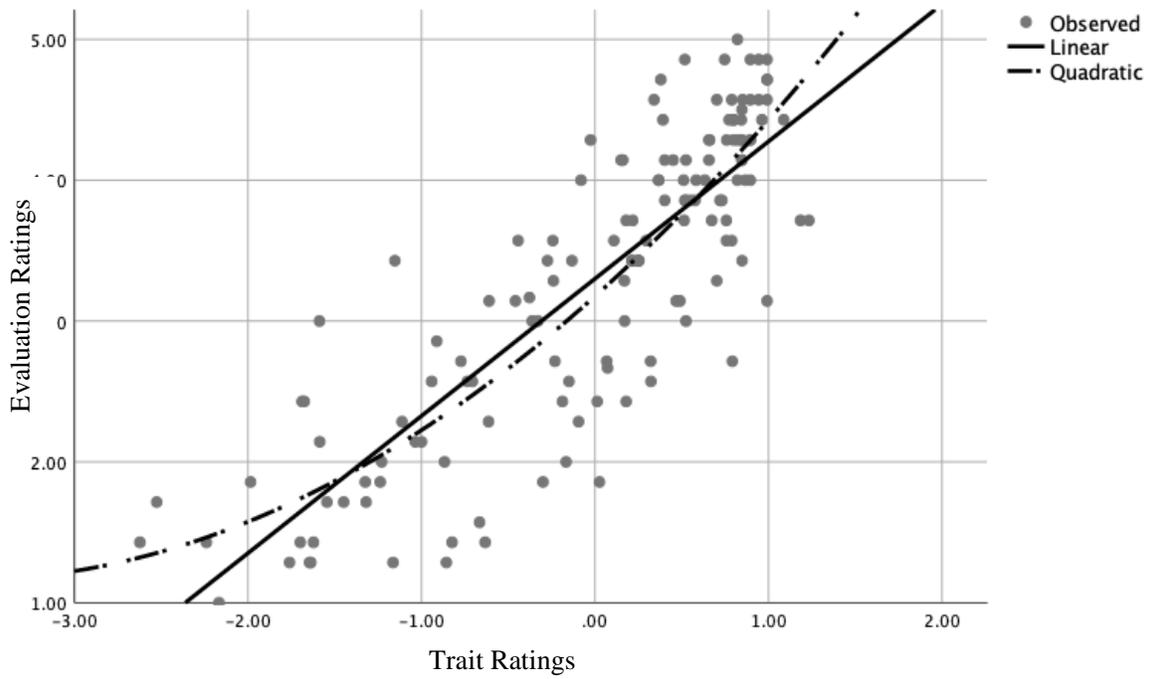


Figure 3. Overall trait and evaluation ratings in male politicians.

In women, there was also a significant convex quadratic relationship between the combined trait variable and the combined evaluation variable, $F(2,142) = 178.85, p < .001$ (see Figure 4). (Note: An examination of the Mahalanobis distance scores indicated one multivariate outlier, but removing this from the analysis did not alter the significance of this relationship).

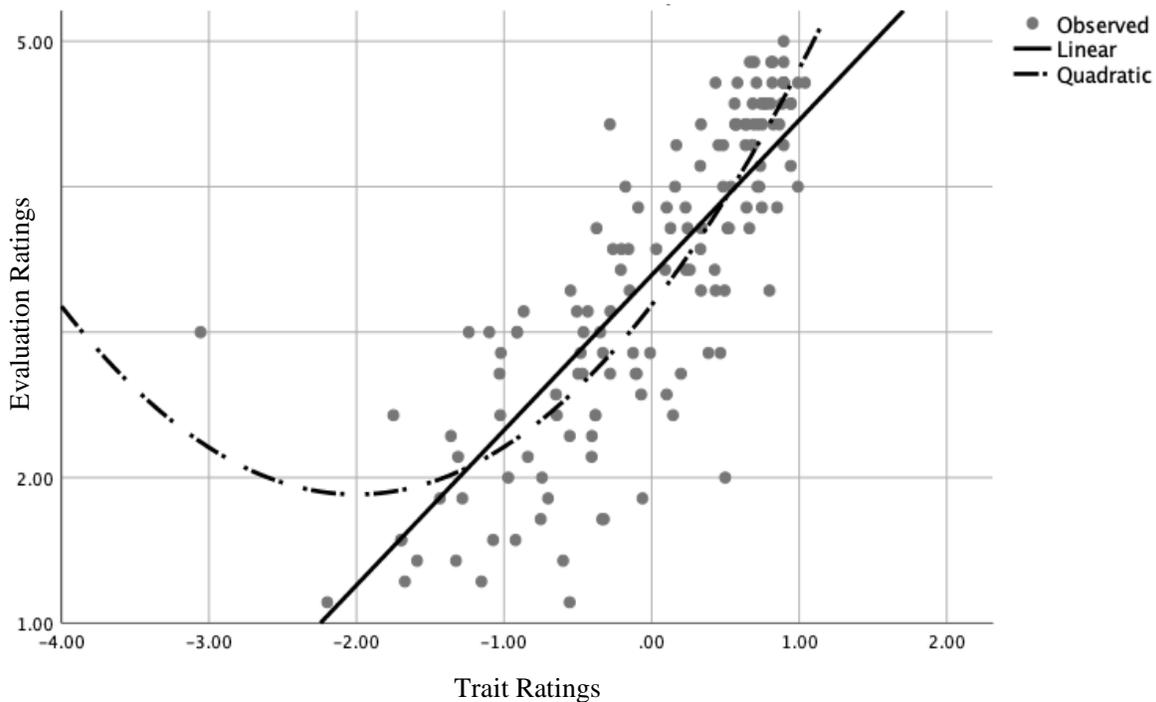


Figure 4. Overall trait and evaluation ratings in female politicians.

As in Study 1, there was no significant impact of any of the exploratory variables (sexism, religiosity, or right-wing authoritarianism) or the gender of the participant. There was also no significant impact of the final three variables: to what degree participants thought women in politics were a) capable, b) likeable and c) electable compared to men.

Discussion

Unlike in Study 1, my hypothesis that women who were deemed warm would be rated significantly more positively than their female colleagues who were not and that as the warmth of female lecturers increased it would relate less strongly to overall ratings was not supported. It was also not the case that the correlation between warmth and overall evaluation of male lecturers strengthened as warmth increases. As a result, my data did not support the idea that women are punished for lacking warmth, while men are valorized when they are warmer than expected.

Interestingly, women were rated as having greater electoral chances in the Democratic primary than men. This result seems discordant given that female candidates received far fewer votes than anticipated in the Democratic primary. For example, Elizabeth Warren – who was considered a frontrunner throughout her campaign, and led national polls for the Democratic nomination in autumn 2019 (Millhiser, 2019) – finished fourth in the New Hampshire primary and the Nevada caucus and fifth in the South Carolina primary, as well as losing her home state of Massachusetts (Goldmacher & Herndon, 2020). This is also at odds with the fact that, when asked about the electability of women in politics generally, respondents indicated that women were slightly less electable than men.

This mis-match as well as the overall lack of replication may be due in some extent to social desirability. This survey was distributed shortly after Joe Biden became the presumptive Democratic nominee, and none of the candidates being assessed were in the race at this time. The widespread emphasis placed on sexism in media coverage of the primaries (e.g., Nilsen & Zhou, 2020; Kurtzleben, 2020; Hunt, 2020) may have caused some respondents to compensate by being overly generous to female candidates whom they did not support during the actual contest.

The concept of electability is perhaps the most major difference between lecturers and politicians, and it makes the perceptions of others relevant to one's own assessment of a politician in a way that does not exist in academia. According to a February 2020 survey conducted by the Public Policy Institute of California, 57% of voters indicated that choosing the presidential candidate most likely to defeat Donald Trump was their most important criteria, while only 33% indicated that selecting a candidate whose policy positions resembled theirs was most important (Baldassare, Bonner, Dykman, & Lawler, 2020). This emphasis on electability may hurt female candidates, as some polls indicate that although the vast majority of Americans

indicate that they themselves would vote for a qualified woman, nearly one-third of them believe their “neighbours” would be unwilling to vote for a woman (Paul & Smith, 2008). The dynamic surrounding electability may allow voters to mask their own prejudice: they can vocally support female candidates while still voting for men under the guise of electability. Previous research supports the notion that people tend to respond in a socially desirable manner when asked about their support for minority groups in politics. Brown-Iannuzzi, Najle, and Gervais (2019) contrasted self-reported “willingness to vote” data with data from indirect tallying measures which allowed some masking of prejudice, and found that 27% of respondents indicated they were unwilling to vote for female political candidates in the indirect measure, compared with only 8% of respondents who indicated the same in the direct poll. Streb, Burrell, Frederick, and Genovese, (2008) used a similar method and found that in indirect measures 26% of respondents were “angry or upset” about the prospect of a female president, a number which was 10 to 20 percent higher than what public opinion polls at the time generally indicated.

It is worth considering whether my research would have benefitted from using indirect measures to avoid these issues related to social desirability. Results may also have been different if this data had been collected before the primary was over when the stakes were higher and I could have asked about actual intentions to vote for a candidate. Indeed, it may also have been beneficial to inquire about which candidate the respondent voted for in the primary (if any) to assess whether the views expressed about the candidate were aligned with actual voting behavior.

Study 3

In order to expand upon the experimental data from Studies 1 and 2, I analyzed existing content for Studies 3 and 4. This study investigated the relative frequency of adjectives related to

warmth and competence in evaluations of male and female lecturers on the popular website “Rate My Professor”, which allows students to anonymously write reviews of their university instructors. To reflect the pattern of punishing women for negative behavior while failing to reward them for positive behavior discussed in previous studies, I hypothesized that negative adjectives would be more frequently used in evaluations of female lecturers, while positive adjectives would be used equally across men and women. I specifically hypothesized that this would be true when evaluating opposite “poles” of the same trait: for example, the positive adjective “organized” would be equally applied to men and women, while the negative adjective “disorganized” would be applied more frequently to women, suggesting enhanced scrutiny of women rather than a true difference in organization levels across gender. I expected that this effect would be stronger for words related to warmth than for words related to competence due to the emphasis placed on warmth in the stereotypical female gender role.

Method

Procedure.

I selected adjectives related to success in teaching according to previous research (Check, 1986; Martinazzi & Samples, 2000). For each positive adjective mentioned, I included its opposite (e.g., fair and unfair) and vice versa. When there was no direct opposite, I selected the closest approximation possible (e.g., knowledgeable and ignorant) using a thesaurus (e.g., Thesaurus.com, n.d.). I entered each word into an online database which assessed the frequency of its use on Rate My Professor according to gender and field of study. If any word had a frequency lower than 1 word per million, it was removed from the list and replaced with a suitable synonym if possible (1 word per million was chosen an arbitrary cut-off point because the gendered language tool presents data in words per million). This resulted in a final word list

containing 84 adjectives: 42 positive and 42 negative; 59 related to warmth (e.g., helpful/unhelpful) and 25 related to competence (e.g., prepared/unprepared). Means and standard deviations were calculated across the 25 fields of study for each word, and these were compared between male and female lecturers.

Measures.

Word Frequency. To calculate the frequency with which different adjectives were used on Rate My Professor, I used the online Gendered Language Tool (Schmidt, 2015). This tool allows users to search for any term, and reports back the frequency with which that term is used per million words on ratemyprofessor.com, categorized according to gender and 25 fields of study (see Figure 5). The tool searches about 14 million reviews from hundreds of colleges and universities, the majority of which are located in the United States.

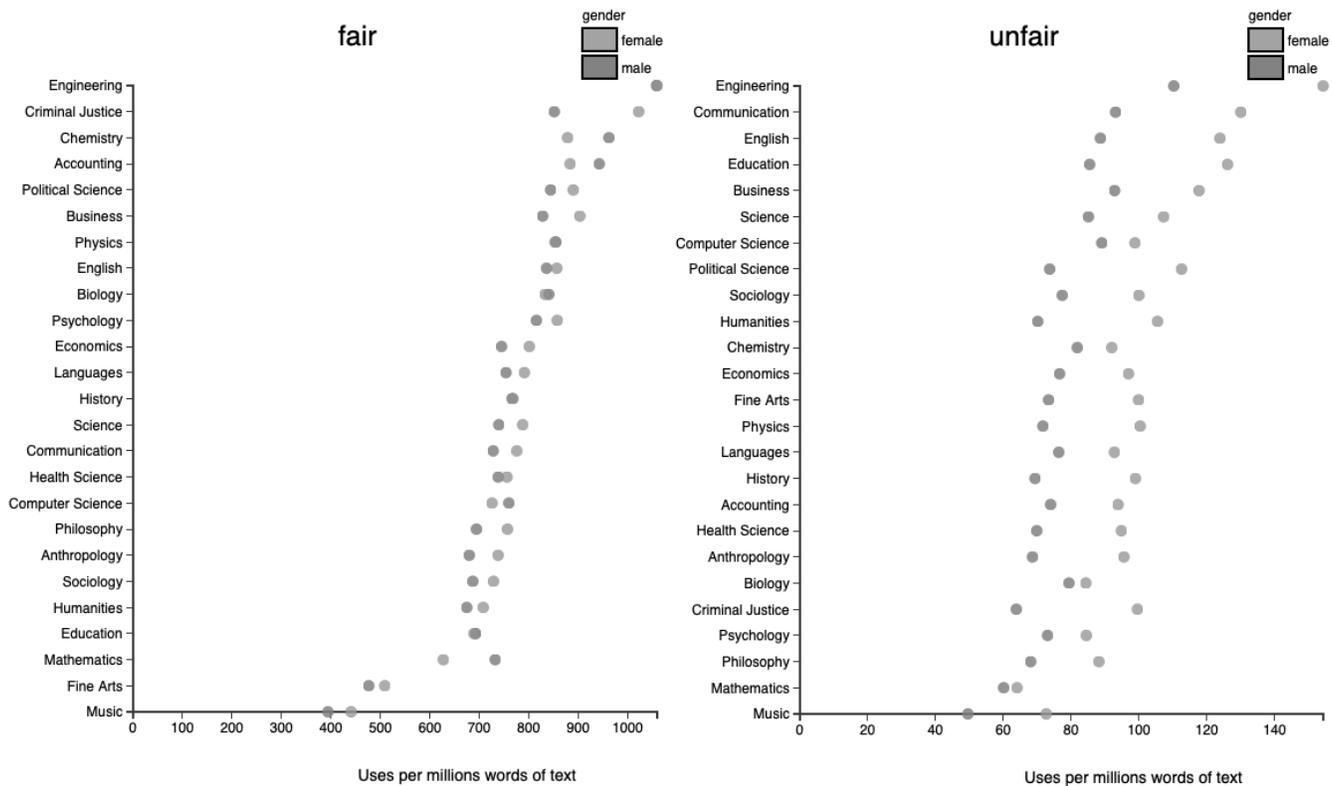


Figure 5. Frequency of use of the words “fair” and “unfair” on Rate My Professor according to the online Gendered Language Tool (Schmidt, 2015).

Results

A 2 x 2 x 2 (Valence x Gender x Adjective Type) ANOVA was conducted on scores, where valence was a within-participants factor indicating the frequency of the positive and negative adjectives associated with each trait; gender was a between-subjects factor indicating the mean word frequency for male and female lecturers on each trait; and adjective type was a between-subjects factor indicating whether the adjective was related to warmth or competence. The results of the mixed-model ANOVA showed that there was a significant main effect of valence, $F(1,80) = 12.56, p = .001, \eta^2 = .136$, such that the positive adjectives associated with each trait ($M = 418.03, SD = 923.70$) were used more frequently than the negative adjectives (M

= 96.19, $SD = 225.72$). However, there was no significant main effect of gender, $F(1,80) = 0.01$, $p = .92$, $\eta^2 = .000$, or adjective type, $F(1,80) = 0.57$, $p = .45$, $\eta^2 = .007$ on word frequency scores. There was also no significant interaction between valence and gender, $F(1,80) = 0.001$, $p = .98$, $\eta^2 = .000$, valence and adjective type, $F(1,80) = 0.44$, $p = .51$, $\eta^2 = .005$, or gender and adjective type, $F(1,80) = 0.01$, $p = .92$, $\eta^2 = .000$. The three-way interaction between valence, gender, and adjective type was also not significant, $F(1,80) = 0.00$, $p = .984$, $\eta^2 = .000$. These findings suggest that there is no significant difference in how positive and negative adjectives are used to describe male and female lecturers across warmth and competence-related traits.

Based on previous research using this database (Storage, Horne, Cimpian, & Leslie, 2016), I also conducted independent sample t-tests comparing male and female lecturers' mean scores for each trait with 48 degrees of freedom based on the number of fields of study. While most mean scores were not significantly different from each other across the two genders, there were 16 traits where my predicted pattern of equally positive ratings for men and women but more negative ratings of women than men did emerge (see Table 1). Ten of these traits were related to warmth (considerate/inconsiderate; kind/unkind; polite/rude; nurturing/unsympathetic; personable/unfriendly; relatable/unapproachable; fair/unfair; calm/crazy; available/unavailable; and honest/fake) and six of these traits were related to competence (prompt/late; prepared/unprepared; competent/incompetent; reliable/unreliable; flexible/rigid; and committed/flaky). There were no instances where the reverse pattern (i.e., equal positive scores across gender but higher negative scores for men than women) emerged.

Table 1

Negative and positive adjective frequency in male and female lecturer evaluations using independent samples t-test for equality of means with 48 degrees of freedom

	Female		Male		t-test
	M	SD	M	SD	
Considerate	19.58	5.53	17.27	4.23	1.66
Inconsiderate	9.78	2.27	7.46	1.73	4.06**
Kind	455.87	80.21	428.24	77.31	1.24
Unkind	2.00	1.23	1.27	0.81	2.48*
Polite	15.24	4.05	13.37	2.83	1.89
Rude	329.73	70.29	224.40	40.03	6.51**
Calm	8.34	3.65	10.05	3.88	1.61
Crazy	164.08	42.55	135.84	24.23	2.88**
Available	231.78	74.76	202.26	61.34	1.53
Unavailable	4.45	1.61	3.49	1.71	2.04*
Nurturing	2.08	2.72	1.02	1.30	1.76
Unsympathetic	2.71	0.99	1.88	0.80	3.26**
Personable	56.83	12.81	57.32	17.12	0.11
Unfriendly	7.87	2.29	4.73	1.66	5.55**
Relatable	4.52	1.74	4.01	1.23	1.20
Unapproachable	22.07	7.07	18.12	5.70	2.17*
Fair	785.39	135.12	763.57	137.31	0.57
Unfair	101.33	18.80	76.80	12.34	5.45**
Honest	59.36	28.58	63.77	35.64	0.48
Fake	14.38	7.64	7.64	2.98	4.11**
Prompt	15.88	7.02	12.69	7.49	1.55
Late	327.74	79.00	256.90	55.89	3.66**
Prepared	336.49	57.82	312.61	52.61	1.53
Unprepared	30.51	8.31	23.03	8.37	3.17**
Competent	13.44	4.99	12.60	5.08	0.59
Incompetent	16.90	7.31	11.17	4.51	3.34**
Reliable	4.52	1.74	4.01	1.23	1.20
Unreliable	3.74	1.27	2.53	1.27	3.37**
Flexible	74.95	31.25	63.75	26.88	1.36
Rigid	8.26	2.74	5.55	1.72	4.19**
Committed	14.68	4.59	13.00	4.92	1.25
Flaky	6.03	4.39	2.02	1.53	4.31**

Note. M = Mean. SD = Standard Deviation. * indicates $p < .05$. ** indicates $p < .01$.

Discussion

I hypothesized that negative words would be more used more frequently with regard to female lecturers, while positive adjectives would be used equally across men and women, even

when the words in question referred to oppositely valenced poles of the same trait. When analyzed in an omnibus fashion using mixed-model ANOVA, this hypothesis was not supported: there was no difference in the frequency with which either positive or negative words were used to describe male and female lecturers for both warmth- and competence-related traits. However, when analyzed on an individual level, a minority of traits did demonstrate the pattern I predicted. Male and female lecturers did not differ on the positive end of the evaluative spectrum, but evaluations of female lecturers more frequently contained words on the negative end of the same spectrum. Notably, there was not a single instance of this pattern where the gender was reversed such that men were the gender receiving disproportionate negative feedback. As I predicted, this pattern more commonly occurred for words related to warmth than for words related to competence, supporting the notion that women are more scrutinized on warmth-related constructs than men are. While these results are not significant on a sample-wide level, they do indicate potential directions for future research. Specifically, research could examine whether there are trait categories more specific than warmth or competence (e.g., sociability) which predict gender differences in evaluation. Another line of future research could build upon the findings of Storage and colleagues (2016), who used this same database to assess the relationship between racial and gender diversity in academic fields and use of the words “brilliant” and “genius”. These authors calculated the frequency with which these words were used in the fields of study mentioned by the Gendered Language Tool and compared that with existing data on the representation of women and African Americans in these fields. A similar tactic could be used in this instance to investigate the frequency of warmth- and competence-related words in more male-dominated (e.g., engineering) and more female-dominated (e.g., sociology) fields.

Study 4

In order to assess the degree to which warmth- and competence-related traits are invoked in media coverage of male and female politicians, I conducted a content analysis of opinion articles from *The New York Times* and *The Washington Post* about candidates for the United States Democratic Party presidential nomination in the lead up to the Democratic Primary. In accordance with previous studies, I hypothesized that warmth-related words would be used more frequently in articles about female candidates than male candidates due to their greater association with traditional female stereotypes. I particularly expected to see negative warmth-related words used more frequently to describe female politicians due to the previously mentioned pattern of punishing women for lacking desirable traits.

Method

Procedure.

I chose to analyze articles from *The New York Times* and *The Washington Post* due to their relatively large readership and influence, as well as their reputations for being standard bearers in the industry. The *Times* has been in print since 1851 and has won 130 Pulitzer Prizes, more than any other newspaper (Peiser, 2020); the *Post* has been in print since 1877 and has won 69 Pulitzer Prizes (“The Washington Post wins”, 2020). A 1999 survey by the *Columbia Journalism Review* ranked the *Times* and the *Post* the first and second best newspapers in the United States, respectively (De Vise, 2011). Using the newspapers’ respective online archives, I searched for articles in the opinion section written about Democratic presidential candidates. I required each article’s date of publication to be after the date the candidate in question announced their presidential campaign and before February 3, 2020, the date of the Iowa Democratic Caucuses. I also required each article to mention the name of the candidate in either

the title or the byline, and not to mention the name of any other candidate in these places. In order to diversify the list of candidates mentioned, I used a maximum of five articles about any given candidate, unless this was not feasible due to a lack of articles about other candidates. This resulted in a list of 64 articles: 32 from the *Times* and 32 from the *Post*. Articles from each publication were divided equally long gender lines. Based on articles available in these archives which met my criteria, I included articles about five female candidates (Elizabeth Warren, Amy Klobuchar, Kamala Harris, Kirsten Gillibrand, and Marianne Williamson) and four male candidates (Joe Biden, Bernie Sanders, Michael Bloomberg, and Pete Buttigieg). I performed a linguistic analysis on these articles which tested for content related to warmth and competence, and compared word frequency in these categories across articles about male and female politicians.

Measures.

Linguistic Content. I analyzed the linguistic content of the articles using the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2007) software. This program computes the frequencies of words being used through the analysis of text data by matching it to a dictionary in which words are assigned to specific categories. I created a dictionary based on a word list by Nicolas, Bai, and Fiske (2019), which contained 364 words categorized into warmth dimensions (sociability and morality) and competence dimensions (ability and agency). This word list also categorized each word by valence as either positive or negative. I also included a second list by these authors which categorized words 375 words into agency and communion dimensions; this word list did not categorize the words by valence.

Results

Contrary to my hypothesis, there were no significant differences between articles about male and female politicians in any of the word frequency categories (see Table 2).

Table 2

Frequency of word use related to warmth and competence dimensions in articles about male and female politicians

		Female		Male		<i>t</i> -test
		M	SD	M	SD	
Warmth	Positive	0.47	0.34	0.42	0.33	0.60
	Negative	0.27	0.29	0.22	0.19	0.82
	Overall	0.73	0.44	0.63	0.42	0.93
Competence	Positive	0.29	0.22	0.27	0.17	0.41
	Negative	0.07	0.11	0.07	0.08	0.00
	Overall	0.36	0.26	0.34	0.19	0.35
Sociability	Positive	0.15	0.16	0.14	0.17	0.24
	Negative	0.02	0.05	0.02	0.05	0.00
	Overall	0.16	0.17	0.16	0.17	0.00
Morality	Positive	0.32	0.28	0.28	0.27	0.59
	Negative	0.25	0.29	0.20	0.20	0.80
	Overall	0.57	0.39	0.48	0.38	0.35
Ability	Positive	0.25	0.18	0.20	0.14	1.23
	Negative	0.02	0.05	0.02	0.07	0.00
	Overall	0.27	0.19	0.23	0.16	0.91
Agency	Positive	0.04	0.09	0.07	0.08	1.39
	Negative	0.05	0.09	0.04	0.06	0.51
	Overall	0.09	0.16	0.11	0.10	0.60
Agency (2)		3.49	1.22	3.28	1.13	0.72
Communion		3.30	1.75	2.78	0.86	1.51

Note. M = Mean. SD = Standard Deviation. * indicates $p < .05$. ** indicates $p < .01$.

However, there was a significant difference between the frequency of agency and communion word use (based on the non-valanced word list) for articles about male, but not female, politicians. For men, agency words ($M = 3.28$, $SD = 1.13$) were used significantly more than communion words ($M = 2.78$, $SD = 0.86$), $t(62) = 2.00$, $p = .05$. For women, there was no significant difference between the frequency with which agency words ($M = 3.49$, $SD = 1.22$) and communion words ($M = 3.30$, $SD = 1.75$) were used, $t(62) = 0.50$, $p = .62$.

Additionally, when correlations were tested between different word categories, significant gender differences emerged. Most notably, the use of positive warmth-related words was associated with the use of negative competence-related words in articles about female, but not male, politicians. The correlation between positive warmth words and negative competence words was significant in women, $r(32) = .63, p < .001$, but not men, $r(32) = -.08, p = .65$, and these correlations were significantly different from each other, $z = 3.12, p = .002$. Likewise, the correlation between positive warmth words and negative agency words was significant in women, $r(32) = .60, p < .001$, but not men, $r(32) = -.06, p = .74$, and these correlations were also significantly different from each other, $z = 2.85, p = .004$.

Discussion

My hypothesis that warmth-related words – particularly negatively valenced ones – would be used more frequently in articles about women than men was not supported. However, there was some evidence for the idea that agency words may be used more frequently to describe male politicians than communion words. Unfortunately, this effect occurred based on dictionary words that were not categorized according to valence, so it is not possible to ascertain whether this effect is being driven by positive or negative words. However, this result may suggest that dimensions related to communion in general – whether positive or negative – are simply not seen as relevant when discussing male politicians.

Interestingly, warmth-related words did have different patterns of correlation with other words in articles about women and men. In articles about female politicians, the use of positive warmth-related words was associated with greater numbers of negative words related to competence; this was not the case in articles about male politicians. This finding aligns with previously mentioned research on the challenges women face balancing the warmth traits

prescribed by the traditional female stereotype with the competence traits prescribed by leadership roles (Rudman & Glick, 1999). While women were not viewed as less competent or less warm overall, the fact that being viewed as more warm corresponded with being viewed as less competent provides some evidence for the idea that displaying traditionally feminine traits may undercut women's ability to be viewed as competent leaders. It is worth noting, however, that the reverse pattern did not emerge: negative warmth-related traits were not correlated with positive competence-related traits, suggesting that women displaying competence are not automatically deemed cold. This conflicts somewhat with Rudman's (1998) backlash hypothesis, which suggests that women face social costs when they present themselves as strong and self-confident as this is perceived as counter-stereotypical behavior. It is possible this is a result of a similar social desirability effect with regards to female politicians as discussed in Study 2: due to the relatively prevalent discussion regarding discrimination against female politicians, perhaps the well-documented pattern of denigrating the warmth of women when they display competence was largely avoided.

Study 5

Due to the unexpected results obtained in Study 4 regarding correlations between warmth and competence words, I attempted a replication to assess whether these effects were robust. This time, I conducted a content analysis of opinion articles from *The Wall Street Journal*, *The Boston Globe* and *CNN* about candidates for the United States Democratic nomination in the lead up to the Democratic Primary. In accordance with my findings from Study 4, I hypothesized that the use of positive warmth-related words would be significantly correlated with the use of negative competence-related words in articles about female politicians, but not male politicians, and that the correlations for men and women would be significantly different from each other.

Method

Procedure.

This time, I chose to analyze articles from *The Wall Street Journal*, *The Boston Globe* and *CNN*, which are all among the most widely read news providers in the United States. The *Journal* has been in print since 1889 and won 37 Pulitzer Prizes (The Wall Street Journal, 2020), and the *Globe* has been in print since 1872 and won 26 Pulitzer Prizes (“Globe numbers look promising”, 2016). *CNN* is primarily a 24-hour cable news channel, but its news website launched in 1995 and is one of the most widely-visited news websites in the world (“CNN”, n.d.). These news providers were chosen over similar competitors based on the accessibility of their opinion coverage of national-level American politics.

I used the same procedure as in Study 4 for locating opinion articles from the websites in question. I compiled a list of 90 articles: 30 from each publication. Half of the articles from each publication were about male candidates (Joe Biden, Bernie Sanders, Pete Buttigieg, and Michael Bloomberg), and half were about female candidates (Elizabeth Warren, Amy Klobuchar, Kamala Harris, Marianne Williamson, and Kirsten Gillibrand). I performed the same linguistic analysis as in Study 4 on these articles which tested for content related to warmth and competence, and compared word frequency in these categories across articles about male and female politicians.

Measures.

Linguistic Content. As in Study 4, I analyzed the linguistic content of the articles using the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2007) software and the same dictionary assessing warmth dimensions and competence dimensions according to valence, as well as agency and communion dimensions (Nicolas et al., 2019).

Results

My hypothesis that positive warmth-related words would be correlated with negative competence-related words in articles about female politicians was not supported. The correlation between positive warmth words and negative competence words was not significant in women, $r(45) = .05, p = .75$, or men, $r(45) = -.17, p = .28$, and these correlations were not significantly different from each other, $z = 1.02, p = .31$. However, in this instance the opposite relationship emerged. The correlation between negative warmth words and positive competence words was significant in women, $r(45) = .39, p = .008$, but not men, $r(45) = -.20, p = .18$, and these correlations were significantly different from each other, $z = 2.83, p = .005$. When assessing the different components of these warmth and competence dimensions, this effect appears to have been driven specifically by the relationship between morality and ability. The correlation between negative morality words and positive ability words was significant in women, $r(45) = .41, p = .005$, but not men, $r(45) = -.17, p = .25$, and these correlations were significantly different from each other, $z = 2.80, p = .005$.

However, in this study there was a significant difference between articles about men and women with regards to communion word use. Contrary to expectations, communion words were used more frequently to describe men ($M = 3.16, SD = 0.84$) than women ($M = 2.75, SD = 1.09$), $t(90) = 2.00, p = .05$.

Discussion

My hypothesis that positive warmth-related words would be associated with negative competence-related words for female politicians (but not male politicians) was not supported. However, the opposite effect emerged for this group in that negative warmth-related words were associated with positive competence-related words. This is more aligned with Rudman's (1998)

backlash hypothesis, in that it demonstrates that women who exhibit dominant, high-status behavior (i.e., high in competence) are perceived as unlikable (i.e., low in warmth). Contrary to expectations, however, this effect appeared to be driven by the relationship between ability and morality, rather than the relationship between agency and sociability, which are the dimensions that align most strongly with masculine and feminine stereotypes (Rudman et al., 2012).

However, Goodwin and colleagues (2014) argue that morality traits may be particularly consistent with the female gender role, as they imply an orientation towards other people and a focus on interpersonal relationships and the community rather than the self.

Surprisingly, the results of this study revealed that more communion-related words were used in articles about men than women, which was somewhat contrary to the findings of Study 4. However, as previously mentioned, this effect was driven by the section of the dictionary that had not been divided into positive and negative words, so it is impossible to know whether this was due to men being praised or punished for communion-related traits. As such, it is hard to make any conclusions based on this effect, other than to note the unexpected finding that communion in general appeared to be more relevant for male than female politicians; further research is needed in this area.

General Discussion

I conducted five studies investigating the role of warmth and competence in the evaluation of women in leadership positions. In Study 1, students evaluated their male and female lecturers on specific traits as well as overall performance; in Study 2 I used this same design to assess the American public's perception of male and female presidential candidates. Study 3 involved the analysis of warmth- and competence-related word frequency in the Rate My Professor reviews of male and female lecturers, while Studies 4 and 5 investigated the

linguistic content of news articles about male and female politicians with regards to warmth- and competence-related words. Overall, the results of these five studies provided mixed support for my overall hypothesis that warmth would be more central to the evaluation of female than male leaders, and specifically that female lecturers and politicians would be punished more for lacking warmth than their male counterparts would be.

In Study 1, when lecturers received low warmth ratings, this rating was much more integral to how female lecturers were evaluated on other dimensions than it was for male lecturers, while at high levels of warmth there was no gender difference. This provided some support for the idea that women are “punished” for lacking warmth traits while men are not, and I set out to replicate this effect in Study 2, replacing lecturers with politicians. However, the results of Study 2 did not align with those of Study 1, suggesting either a fundamental difference between lecturers and politicians or the lack of a robust effect in the initial study. When returning to lecturers in Study 3 vis-à-vis Rate My Professor evaluations, I found some support for the greater scrutiny of women’s warmth in that references to several warmth-related traits followed a pattern where reviews of men and women did not differ on the positive end of the trait spectrum, but women were more likely than men to be referred to with words on the negative end of the same spectrum. This result tied in with the findings from Study 1, once again suggesting that women may be more likely than men to be punished for lacking warmth (but not necessarily praised for possessing it). Studies 4 and 5 were a return to analyzing politicians, this time through the lens of news media. Neither of these studies provided direct evidence of my hypothesis (i.e., through the inclusion of more negative warmth-related content in articles about female politicians), but there was some evidence of an opposing relationship between warmth and competence for women. This could be seen in the correlations between positive warmth

content and negative competence content (Study 4) and between negative warmth content and positive competence content (Study 5) that existed for female politicians but not male politicians.

Lecturers vs. Politicians

To some extent, my studies on lecturers (Study 1 and Study 3) provided greater support for my hypothesis than my studies on politicians (Studies 2, 4 and 5). As previously discussed, this may have been related to social desirability and the way in which electability can serve as a means to mask one's own prejudice (Paul & Smith, 2008). Politicians attain their positions through gathering popular support, and thus their appeal to people beyond one's self is relevant in their evaluation; this is not the case with lecturers. There is also an argument to be made that, partially due to the previously mentioned widespread media coverage of sexism in politics (e.g., Nilsen & Zhou, 2020; Kurtzleben, 2020; Hunt, 2020), gender was simply more salient to participants in the politician studies than the lecturer studies. This is especially worth considering given that the politicians in question were vying for the opportunity to be elected president of the United States: a position that has never been held by a woman. Conversely, students within the School of Psychology at the University of Kent were likely relatively accustomed to being taught by female lecturers. This may have rendered their gender less salient and reduced the tendency towards evaluating women in an artificially positive manner in order to conform to social norms regarding gender equality (Streb et al., 2008).

However, it is worth noting that there may be other differences between lecturers and politicians that result in differential emphasis placed on warmth and competence traits. For example, students are much more likely to interact in a one-on-one manner with their lecturers than citizens are with their political representatives, and this interpersonal interaction may mean that the prescription for feminine warmth is more intense for lecturers than politicians. Indeed,

El-Alayli and colleagues (2018) found that perceptions of women as more communal manifested in students expecting more emotional support, asking for more special favours, and initiating more friendship behaviours with female lecturers than male lecturers; a result that is unlikely to be mirrored in politicians due to the lack of opportunity for one-on-one-interactions. The construction of both the female stereotype and the dimension of warmth is very related to interpersonal interaction: characteristics such as being nurturing, sympathetic, and helpful are much easier to enact in a one-on-one setting as opposed to a in a public-facing role. As a result, this may lead to more opportunities for lecturers than politicians to be perceived as violating or fulfilling their gendered expectations regarding warmth-related behaviours.

Some of the discrepancy between the Study 1 and Study 2 findings may also have arisen due to the different participant samples used. Study 1 surveyed British university students, while Study 2 relied on American users of Prolific Academic. Participants in Study 1 were more likely to be female (85.75%, vs. 56.44% in Study 2), and their mean age was nearly 15 years younger than participants in Study 2 (19.30 years vs. 34.07 years). Past research has also uncovered some personality differences between university samples and crowdsourced samples such as those from Prolific Academic. For example, Colman, Vineyard and Letzring (2018) found that university student participants were lower in openness, and higher in extraversion, agreeableness, and neuroticism than participants from the crowdsourcing website MTurk. Another notable difference between student samples and crowdsourced samples such as those on Prolific Academic is the greater experience crowdsourced participants accumulate with social science studies. These individuals (typically) spend a large amount of time completing studies like this one, and as a result may be less naïve participants (Hauser, Paolacci, & Chandler, 2018). It is worth considering whether this may have resulted in, for instance, participants correctly

deducing that the aim of this study was to assess effects related to the politician in question's gender, despite it not being explicitly stated.

Potential Impact

Despite only partial support my hypothesis, the potential ramifications of women being evaluated based on different criteria than their male counterparts with regards to warmth-related traits could be significant. In academia, decisions regarding hiring, tenure and promotion are still largely influenced by scores on students teaching evaluations (Mengel et al., 2018), so bias in evaluations may result these processes being unintentionally biased against female lecturers. Apart from the negative impact on the lecturer in question, this can have further downstream effects on students: for instance, research shows that in male-dominated science, technology, engineering and math (STEM) classes, female students perform better when their instructor is a woman (Boring, 2017). With this in mind, negative evaluations of female lecturers, may result in a self-perpetuating cycle where women continue to be underrepresented in academia.

Although my main hypothesis was not supported within the domain of politics, my findings regarding the different correlations warmth-related words have with other positive and negative traits in articles about male and female politicians still hints at warmth playing a different role for these groups. Dolan and Hansen (2018) claim that women face hidden barriers to being elected to political office, and this may constitute one of them. Given that increases in the number of female representatives has been linked to greater funding for social services such as public health (Clayton & Zetterberg, 2018), the implications of gender bias in the electoral process may have significant consequences for society as a whole.

Previous research has also suggested that media coverage plays a prominent role in political sexism. According to the 2009 European Election Study's Media Content Data (Lühiste

& Banducci, 2016), women candidates appear less frequently overall in the media than their male counterparts: this was certainly true in the sources I used, in which I largely struggled to find suitable articles about female candidates apart from Elizabeth Warren, while a number of male candidates were frequently written about in each publication. Haraldsson and Wängnerud (2019) found that, at the aggregate level, the higher the level of sexism in the media in a given country, the lower the share of women candidates in that country, suggesting a cycle where media sexism may both reflect sexism in society at large and perpetuate it. It is worth noting that I used predominantly center- or left-leaning media institutions for my content analyses, with the possible exception of *WSJ* whose editorial pages lean slightly conservative (“The Wall Street Journal”, n.d.). Given the emphasis on egalitarianism in liberal institutions, it is worth considering whether I may have observed findings more in line with my hypotheses had I sampled papers from across the political spectrum. Previous research has shown that endorsing complementary stereotypes of men as agentic (but not communal) and women as communal (but not agentic) is linked to system justification, a tendency linked to conservatism (Jost & Kay, 2005). Additionally, Blumell (2018) found that online political reporters working for conservative websites endorsed higher levels of hostile sexism than those working for liberal websites. Hostile sexism involves punishing women for acting in stereotypically unfeminine ways (Glick & Fiske), such as competing against men in the political realm – as a result, Blumell (2018) suggested that this hostility might negatively impact coverage of women in politics, and lead to enhanced scrutiny. Given these findings, I believe it is likely that articles from conservative-leaning news outlets would have adhered more strongly to gendered stereotypes, and women would have been more scrutinized along warmth-related lines than they were in liberal media outlets. While I felt that the reality of the 2020 primary only occurring in the

Democratic party (due to President Donald Trump fulfilling the role of the Republican party incumbent) necessitated choosing newspapers from the side of the political spectrum associated with the primary (i.e., liberal), future research could endeavor to compare coverage of female politicians in liberal and conservative media outlets (taking into account the relative scarcity of right-of-center sources deemed “reliable”, according to Ad Fontes Media [2020]). While it would undoubtedly also be interesting to examine female politicians from the Republican party, this would unfortunately be rendered difficult by their small numbers – of the 17 candidates who entered the race for the Republican presidential nomination in the 2016 election, only one (Carly Fiorina) was a woman (“2016 Republican Party presidential primaries”, n.d.).

Limitations

This research had several limitations. As previously mentioned, the impact of social desirability in answering questions related to sexism must be taken into account. Given that gender bias in politics and academia has gained public attention in recent years, social desirability could have been mitigated in my experimental studies (Study 1 and Study 2) by using indirect rather than direct measures. Previous research (Brown-Iannuzzi et al., 2019; Streb et al., 2008) has succeeded in lessening socially desirable responses using the “unmatched count technique” rather than self-report measures. This technique involves participants in the control condition seeing a list of innocuous statements (e.g., “I own a dog”, “I enjoy pizza”), while participants in the experimental condition see the same list of statements plus the sensitive statement of interest (e.g., “I have cheated on my spouse”). All participants are asked to merely tally the *number* of statements that apply to them, rather than indicating *which* statements apply to them. Given that participants in the two conditions see the same list of items apart from the item of interest, researchers can infer that any difference in the two groups is due to that item.

This technique has been proven to both increase the proportion of respondents willing to endorse socially undesirable items (e.g., “I do not think women should be in politics”) and decrease the proportion of respondents who endorse socially desirable items (e.g., “I volunteer on a regular basis”). My research may have benefitted from using a technique like this given the sensitive subject matter, as self-report measures rely on participants being willing to disclose their socially inappropriate attitudes.

A second limitation with regards to measures in the two experimental studies is that I only asked about the participants perceptions of the traits the lecturers and politicians possessed, not about specific behaviours. For example, MacNeill and colleagues (2015) had online instructors all post students’ grades two days after submission, and found that instructors perceived as male were rated higher on a promptness scale than their colleagues perceived as female. The fact that these researchers compared ratings based on a concrete behavior (the number of days before feedback was given) allows for an absolute comparison across genders because the behavior can be tightly controlled; in my study, I have no way of knowing if, for example, female lecturers are simply less warm on average than male lecturers, so it is more difficult to tease out which effects occur solely due to gender bias. This could have been at least partially avoided by asking students more specific questions about their lecturer’s behavior, such as “how many times has this lecturer assisted you outside of class time”, or “how long does it take this lecturer to provide feedback on your assignments”, rather than only asking for the students’ perceptions of their lecturers’ traits.

Another limitation was the relatively small sample of targets (University of Kent psychology lecturers and Democratic presidential candidates) in all studies apart from Study 3. The design of these studies did not easily allow for larger pools of targets, but the fact that they

are few in number makes it more likely that any effects could be swayed by the individual-level attributes of one or more lecturers/politicians, rendering it more difficult to generalize about these findings. Other attributes of these target groups make generalizability more limited as well: Study 1 only used lecturers from the School of Psychology, which is a relatively female-dominated social science discipline (Mantle, 2020). Results may have been different if I had used lecturers from one of the more male-dominated STEM disciplines. Similarly, Study 2 only included politicians from the Democratic party, which in the modern era is widely regarded as the more liberal and egalitarian of the two major American political parties (Carr, Gamm, & Phillips, 2016). My findings may have skewed differently had I looked at politicians from the Republican party; unfortunately, since the upcoming 2020 election involves an incumbent Republican president, this was not possible at this time using presidential candidates.

While Study 3 involved the use of a database composed of millions of lecturer reviews and thus did not suffer from the same sample size issues, this study had its own limitations. Namely, since I did not have access to the original evaluations, it was impossible to know in what context a given adjective was used. For example, a positive adjective could be used with a negative qualifier (e.g., “smart” vs. “not smart”), or an adjective could be referring to coursework rather than the lecturer in question. Additionally, many adjectives did not necessarily have a direct opposite but rather several synonymous opposites, making it difficult to ascertain whether the words I chose represented the full trait spectrum accurately.

When it comes to my archival studies assessing news media (Studies 4 and 5), there was a slightly different set of limitations. As previously mentioned, it was more difficult to find articles within my criteria (opinion pieces with one candidate only mentioned in the title and/or byline, published in the timeframe between the candidate’s running announcement and the Iowa

caucus) that were written about female politicians as compared to male politicians. The one exception to this rule was Massachusetts Senator Elizabeth Warren, who was frequently written about in all five publications sampled. Of the 77 articles about female candidates that I included in my analysis, 31 of them were about Elizabeth Warren, constituting almost half the sample. Given that no one male candidate comprised such a large portion of the article subjects, this may have resulted in a somewhat lopsided analysis where effects for female politicians were driven too strongly by individual attributes of Senator Warren. Additionally, as previously mentioned, the news publications I chose predominantly skew to the liberal side of the political spectrum. Ad Fontes Media is a company composed of news analysts across the political spectrum who systematically rate the content of news articles (Ad Fontes Media, 2020); this company reviews articles in terms of political bias and uses the weighted average of these scores to assign news sources an overall bias rating on a scale of -42 to +42, with higher negative scores being more left, higher positive scores being more right, and scores close to 0 being the closest to politically neutral. *CNN* was given a bias rating of -5.69, *The New York Times* was given a rating of -4.01, *The Washington Post* was given a rating of -4.18, and *The Wall Street Journal* was given a rating of 1.89. This indicates that the first three publications all lean slightly left, while the *WSJ* leans center-right. *The Boston Globe* was not analyzed by Ad Fontes Media, but editors describe the *Globe* as a liberal institution (Buccini, 2001). As a result, these news sources may not accurately represent the range of views that exist across the political spectrum regarding the politicians in question.

Future Research

This research opens up many possible avenues for future research. To begin with, research could be conducted that addresses some of the limitations mentioned above. For

instance, experimental research conducted using both indirect measures that help mitigate socially desirable responding as well as measures that ask participants about specific, measurable behaviours would be beneficial to highlight some effects that may not be visible in my current research. Additionally, research that widens the demographics beyond female-dominated (and liberal) psychology departments as well as liberal politicians and news outlets would be welcome, particularly given the tendency of social science research to only sample a limited portion of the population that is not representative of society at large. Given the aforementioned relationship between sexism and conservatism, future research would benefit from testing for these effects in a conservative sample. The limitation regarding the impact of individual-level attributes of the lecturers and politicians chosen could also be overcome by replacing the real-world examples used with generic lecturers and politicians; this would also make it possible to manipulate levels of warmth and competence.

Future research could also be conducted that expands on some of the findings I uncovered. For example, in Study 1 I found that women were more punished for warmth deficits than men were; a future study could delve further into the process behind this gender bias, investigating the motivations behind it and whether they align with previous theories such as Rudman's (1998) backlash hypothesis. Additionally, in Study 3 I found some evidence for the idea that female lecturers were disproportionately described using negative adjectives, while descriptions using positive adjectives did not differ between male and female lecturers. Given that this effect was found using archival research, it would be beneficial to expand upon this by conducting an experimental study to ascertain whether an effect is truly present.

Future research could also look at other groups of high-status women beyond lecturers and politicians. For example, previous research suggests that similar gender bias exists with

regards to business professionals such as managers and CEOs (Heilman & Okimoto, 2007), suggesting this is another realm where the role of warmth and competence could be examined. Looking at politicians in countries beyond the United States could also be beneficial, given that cultural differences exist with regards to how politicians are viewed; indeed, examining countries that currently have a female leader (e.g., New Zealand, Germany) could help uncover perceptions of female heads of state when they are already performing their role, rather than just in the electoral process.

Conclusion

Although I found some evidence that warmth is more central to the evaluation of women in leadership than men, and specifically that women are more likely than men to be perceived negatively when they lack warmth, the evidence is far from conclusive. However, my findings related to women being punished more than men for lacking warmth-related traits have wide-ranging potential impacts for women in politics and academia as well as other leadership positions. My findings also open up possibilities for various avenues of future research into evaluations of high-status women.

References

- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in psychology, 7*, 1810. <https://doi.org/10.3389/fpsyg.2016.01810>
- Ad Fontes Media. (2020). Retrieved from <https://www.adfontesmedia.com>
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Baldassare, M., Bonner, D., Dykman, A., & Lawler, R. (2020). PPIC Statewide Survey: Californians and Their Government. *Public Policy Institute of California*. Retrieved from <https://www.ppic.org/publication/ppic-statewide-survey-californians-and-their-government-february-2020/>
- Bligh, M. C., Schlehofer, M. M., Casad, B. J., & Gaffney, A. M. (2012). Competent enough, but would you vote for her? Gender stereotypes and media influences on perceptions of women politicians. *Journal of Applied Social Psychology, 42*, 560-597. Doi: 10.1111/j.1559-1816.2011.00781.x
- Blumell, L. E. (2018). Bro, foe, or ally? Measuring ambivalent sexism in political online reporters. *Feminist Media Studies, 1-17*. Doi: 10.1080/14680777.2018.1546211
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of public economics, 145*, 27-41. <http://dx.doi.org/10.1016/j.jpubeco.2016.11.006>
- Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135–143. doi:10.1002/ejsp.744

- Brown-Iannuzzi, J. L., Najle, M. B., & Gervais, W. M. (2019). The illusion of political tolerance: Social desirability and self-reported voting preferences. *Social Psychological and Personality Science, 10*, 364-373. <https://doi-org.chain.kent.ac.uk/10.1177/1948550618760147>
- Buccini, Cynthia K. (2001). Every Day Is Judgment Day. *Bostonia*. Retrieved from <https://web.archive.org/web/20111107042610/http://www.bu.edu/bostonia/fall01/loth/>
- Burn, S. M., & Busso, J. (2005). Ambivalent sexism, scriptural literalism, and religiosity. *Psychology of Women Quarterly, 29*, 412–418. Doi:10.1111/j.1471-6402.2005.00241.x
- Carr, M., Gamm, G., & Phillips, J. (2016). Origins of the Culture War: Social Issues in State Party Platforms, 1960–2014. *Paper resented at the American Political Science Association*. <http://www.columbia.edu>.
- Chalmers, J. K. (2019, November 6). Evaluations of Lecturers. Retrieved from osf.io/2fnhs
- Chalmers, J. K. (2020, March 4). Evaluations of Lecturers. Retrieved from osf.io/wsp3r
- Check, J. F. (1986). Positive traits of the effective teacher—negative traits of the ineffective one. *Education, 106*, 326 – 334.
- Clayton, A., Zetterberg, P. (2018). Quota Shocks: Electoral Gender Quotas and Government Spending Priorities Worldwide. *Journal of Politics, 80*, 916-932. <https://doi.org/10.1086/697251>
- Colman, D. E., Vineyard, J., & Letzring, T. D. (2018). Exploring beyond simple demographic variables: Differences between traditional laboratory samples and crowdsourced online samples on the Big Five personality traits. *Personality and Individual Differences, 133*, 41-46. <https://doi.org/10.1016/j.paid.2017.06.023>
- CNN. (n.d.) In *Wikipedia*. Retrieved June 20, 2020 from <https://en.wikipedia.org/wiki/CNN>

- De Vise, D. (2011, October 4). What if the rankers ranked newspapers? *The Washington Post*. Retrieved from https://www.washingtonpost.com/blogs/college-inc/post/what-if-the-rankers-ranked-newspapers/2011/10/04/gIQAYZl6KL_blog.html
- Dolan, K., & Hansen, M. (2018). Blaming Women or Blaming the System? Public Perceptions of Women's Underrepresentation in Elected Office. *Political Research Quarterly*, 71, 668-680. <https://doi.org/10.1177/1065912918755972>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3), 573. Doi: 10.1037//0033-295X.109.3.573
- El-Alayli, A., Hansen-Brown, A. A., & Ceynar, M. (2018). Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles*, 79, 136-150. <https://doi.org/10.1007/s11199-017-0872-6>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11, 77-83. Doi: 10.1016/j.tics.2006.11.005
- Fox, R. L., & Lawless, J. L. (2011). Gendered perceptions and political candidacies: A central barrier to women's equality in electoral politics. *American Journal of Political Science*, 55, 59-73. <https://doi-org.chain.kent.ac.uk/10.1111/0022-4537.00128>
- Fulton, S. (2012) Running Backwards and in High Heels: The Gendered Quality Gap and Incumbent Electoral Success. *Political Research Quarterly*, 65, 303-314. <https://doi.org/10.1177/1065912911401419>
- Garin O. (2014) Ceiling Effect. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_296

- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70, 491–512.
doi:10.1037/0022-3514.70.3.491
- Globe numbers look promising. (2016, October 3). *CommonWealth Magazine*. Retrieved from <https://commonwealthmagazine.org/the-download/globe-numbers-look-promising/>
- Goldmacher, S., & Herndon, A. (2020, March 5). Elizabeth Warren, Once a Front-Runner, Drops Out of Presidential Race. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/03/05/us/politics/elizabeth-warren-drops-out.html>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of personality and social psychology*, 106, 148. Doi: 10.1037/a0034726
- Haraldsson, A., & Wängnerud, L. (2019). The effect of media sexism on women's political ambition: evidence from a worldwide study. *Feminist media studies*, 19, 525-541. Doi: 10.1080/14680777.2018.1468797
- Hauser, D., Paolacci, G., & Chandler, J. J. (2018). Common concerns with MTurk as a participant pool: Evidence and solutions. Doi: 10.31234/osf.io/uq45c
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of social issues*, 57, 657-674. <https://doi.org/10.1111/0022-4537.00234>
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: the implied communality deficit. *Journal of applied psychology*, 92, 81-92. Doi: 10.1037/0021-9010.92.1.81

- Hunt, S. (2020, March 11). Four women were serious candidates for president. What happened? *CNN*. Retrieved from <https://www.cnn.com/2020/03/06/opinions/sexism-2020-election-opinion-hunt/index.html>
- Imhoff, R., & Koch, A. (2017). How orthogonal are the Big Two of social perception? On the curvilinear relation between agency and communion. *Perspectives on Psychological Science, 12*, 122-137. Doi: 10.1177/1745691616657334
- Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *Journal of personality and social psychology, 88*, 498. Doi: 10.1037/0022-3514.88.3.498
- Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology, 48*, 77-85. Doi:10.1016/j.jesp.2011.08.001
- Kurtzleben, D. (2020, April 17). Did Gender Keep Democratic Women From Winning The Presidential Primary? *NPR*. Retrieved from <https://www.npr.org/2020/04/17/818952460/did-gender-keep-democratic-women-from-winning-the-primary>
- Levesque, R. J. (2011). Sex roles and gender roles. *Encyclopedia of adolescence, 2622-2623*. <https://doi.org/10.1007/978-1-4419-1695-2>
- Lühiste, M., & Banducci, S. (2016). Invisible Women? Comparing Candidates' News Coverage in Europe. *Politics & Gender, 12*, 223–253.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 291-303. <https://doi-org.chain.kent.ac.uk/10.1007/s10755-014-9313-4>

- Manning, J., & Brudnick, I. (2020). Women in Congress: Statistics and Brief Overview. Congressional Research Service. Retrieved from <https://fas.org/sgp/crs/misc/R43244.pdf>
- Mantle, R. (2020, January 29). Higher Education Student Statistics: UK, 2018/19 - Subjects studied. Retrieved from <https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics/subjects>
- Martinazzi, R., & Samples, J. (2000, October). Characteristics and traits of an effective professor. In *30th Annual Frontiers in Education Conference. Building on A Century of Progress in Engineering Education. Conference Proceedings (IEEE Cat. No. 00CH37135)* (Vol. 2, pp. F3F-7). IEEE.
- Mengel, F., Sauermann, J., & Zölitz, U. (2018). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17, 535-566. [https://doi-org.chain.kent.ac.uk/10.1093/jeea/jvx057](https://doi.org/chain.kent.ac.uk/10.1093/jeea/jvx057)
- Millhiser, I. (2019, October 8). Elizabeth Warren is now leading the 2020 polls. *Vox*. Retrieved from <https://www.vox.com/2019/10/8/20905274/elizabeth-warren-frontrunner-democratic-nomination-2020>
- Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51, 648-652. Doi: <https://doi.org/10.1017/S104909651800001X>
- Nicolas, G., Bai, X., & Fiske, S. T. (2019). Automated Dictionary Creation for Analyzing Text: An Illustration from Stereotype Content. <https://doi.org/10.31234/osf.io/afm8k>
- Nilsen, E., & Zhou, L. (2020, March 6). Why women are feeling so defeated after Elizabeth Warren's loss. *Vox*. Retrieved from <https://www.vox.com/2020/3/6/21166338/elizabeth-warren-loss-2020-primary-sexism>

- Paul, D., & Smith, J. L. (2008). Subtle sexism? Examining vote preferences when women run against men for the presidency. *Journal of Women, Politics & Policy*, 29, 451-476.
Doi: 10.1080/15544770802092576
- Pennebaker, J. W., Francis, M. F., & Booth, R. J. (2007). Linguistic Inquiry and Word Count (LIWC): LIWC2007. Mahwah, NJ: Lawrence Erlbaum Associates.
- Peiser, J. (2020, May 4). The New York Times Won 3 Pulitzers. Read the Winning Work. *The New York Times*. Retrieved from
<https://www.nytimes.com/2020/05/04/business/media/new-york-times-pulitzer-prizes.html>
- Prentice, D. A., & Carranza, E. (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, 269-28. <https://doi.org/10.1111/1471-6402.t01-1-00066>
- Rakow, L. F. (1991). Gender and race in the classroom: Teaching way out of line. *Feminist Teacher*, 6, 10–13.
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, 74, 629. <https://doi.org/10.1037/0022-3514.74.3.629>
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, 77, 1004.
<http://dx.doi.org/10.1037/0022-3514.77.5.1004>

- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues, 57*, 743-762. <https://doi.org/10.1111/0022-4537.00239>
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48*, 165-179. <https://doi.org/10.1016/j.jesp.2011.10.008>
- Schmidt, B. (2015) Gendered Language in Teaching Reviews. Available from <http://benschmidt.org/profGender/>
- Sibley, C. G., Wilson, M. S., & Duckitt, J. (2007). Antecedents of men's hostile and benevolent sexism: The dual roles of social dominance orientation and right-wing authoritarianism. *Personality and Social Psychology Bulletin, 33*, 160–172.
- Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles, 53*, 779-793. Doi: 10.1007/s11199-005-8292-4
- Storage, D., Horne, Z., Cimpian, A., & Leslie, S. J. (2016). The frequency of “brilliant” and “genius” in teaching evaluations predicts the representation of women and African Americans across fields. *PloS one, 11*. doi: 10.1371/journal.pone.0150194
- Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly, 72*, 76-89. doi:10.1093/poq/nfm035
- Sullivan, K. T. (2001). Understanding the relationship between religiosity and marriage: An investigation of the immediate and longitudinal effects of religiosity on newlywed couples. *Journal of family psychology, 15*, 610. <http://dx.doi.org/10.1037/0893-3200.15.4.610>

Thesaurus.com (n.d.). Retrieved April 27, 2021 from <https://www.thesaurus.com>

The Wall Street Journal (2020). Retrieved June 20, 2020 from

<https://www.dowjones.com/products/wsj/>

The Wall Street Journal. (n.d.) In *Wikipedia*. Retrieved July 4, 2020 from

https://en.wikipedia.org/wiki/The_Wall_Street_Journal

The Washington Post wins the 2020 Pulitzer Prize for Explanatory Reporting for groundbreaking climate change coverage. (2020, May 4). *The Washington Post*. Retrieved from

<https://www.washingtonpost.com/pr/2020/05/04/washington-post-wins-2020-pulitzer-prize-explanatory-reporting-groundbreaking-climate-change-coverage/>

2016 Republican Party presidential primaries. (n.d.) In *Wikipedia*. Retrieved April 29, 2021 from

https://en.wikipedia.org/wiki/2016_Republican_Party_presidential_primaries

Żemojtel-Piotrowska, M. A., Marganski, A., Baran, T., & Piotrowski, J. (2016). Corruption and Sexual Scandal: The Importance of Politician Gender. *Anales De Psicología / Annals of Psychology*, 33, 133-141. <https://doi.org/10.6018/analesps.32.3.229171>