

## Research Article

Celina Stolz\*, Marc-André Reinhard, Luise Ende

# Mean girls, queen bees and iron maidens? Female leadership and accusations of workplace bullying

<https://doi.org/10.1515/psych-2022-0127>

received July 1, 2021; accepted May 18, 2022.

**Abstract:** We examined whether female leaders would be evaluated less favorably compared to male leaders regarding workplace bullying. Previous research has demonstrated that women violating prescriptive gender norms of communality experience backlash, and that female leaders are stereotyped of having a *communality deficit*. Building on that, we hypothesized (1) more moral outrage against and (2) more intentions to punish a female leader compared to a male leader. We further hypothesized (3) the accusations of workplace bullying against a female leader were going to be judged as more accurate than against a male leader. Further, defendants that stereotypically fit to the crime they are accused of were found to be judged guilty more often. So, we assumed, (4) a suspected bully that is a female leader was going to be judged as less credible, while (2) the suspected victim of a female leader bully was going to be judged as more credible compared to a male leader. Participants ( $N = 202$ ) read a workplace bullying scenario with a female employee accusing either a female or a male leader of bullying. No effect of gender of suspected bully was found for moral outrage measures, punishment intention judgments, and credibility judgments. Contrary to our predictions, participants found the accusations against the male leader significantly more accurate than against the female leader. Gender and sex-role scores of participants were found to be linked to judgments. Implications for future research are discussed.

**Keywords:** female leadership; workplace bullying; gender stereotypes; communality deficit; backlash; credibility judgments; moral outrage.

## 1 Theoretical Background

Do we judge women, especially successful women, by higher moral standards than men? (Rubner, 2020). Addressing this, as an example, the case of Heike Egner, professor at the University of Klagenfurt is mentioned. She was dismissed from her position after being accused of workplace bullying by one of her postgraduate students. Her former employee stated that she was unfairly and severely criticized by Egner. While Egner admitted that there was a conflict, she described her own behavior as constructive criticism. This textbook example of one word against another lost Egner her job. Persons in leadership positions in academia getting accused of mistreating their employees seems to be a phenomenon asymmetrically often happening to *female* professors, as women are still underrepresented in academic leadership positions. The *Die Zeit* article by Rubner (2020) cites an internal, unpublished study of a college, the Eidgenössischen Polytechnischen Hochschule Lausanne (EPFL). It was found that while only 15 percent of professors at the college were female, 24 percent of the female professors in their study had been accused of workplace bullying – why? Are female bosses actually meaner and colder compared to male bosses? Or are they judged differently than men, and so accused more often? Research on gender stereotypes suggests that women in leadership positions are still confronted with very different expectations than men and therefore are perceived and consequently treated differently (Burgess & Borgida, 1999; Heilman et al., 2004).

\*Corresponding author: Celina Stolz, Department of Psychology, University of Kassel, Germany, E-mail: celina.stolz@web.de  
Marc-André Reinhard, Luise Ende, Department of Psychology, University of Kassel, Germany

## 2 Communality and Agentic Stereotypes

Descriptive gender stereotypes are widespread beliefs about how men and women are. While men are typically seen as agentic, meaning assertive, women are described as communal, meaning sensitive (Eagly & Mladinic, 1989; Hentschel et al., 2019; Prentice & Carranza, 2002). Across 30 nations, men were continually rated as being more agentic while women were rated as being more communal (Williams & Best, 1990). Additionally, the prescriptive component of gender stereotypes assigns how men and women “should be” (Burgess & Borgida, 1999; Eagly, 1987). For the construction of the widely used sex-role inventory *BSRI*, participants were asked to indicate which characteristics are socially desirable for men and which for women. Socially desirable characteristics for men were - amongst others - acting as a leader, being dominant, assertive and forceful. For women, it was rated as socially desirable to be gentle, affectionate, sensitive to the needs of others, warm and understanding (Bem, 1974; Koenig, 2018).

In line with the communality gender norm, women are significantly more altruistic, and are expected to be so (Brañas-Garza et al., 2018; Innocenti & Paziienza, 2006). Heilman and Chen (2005) found that offering help to a co-worker lead to positive evaluations of men but had little effect on the evaluation of women. In contrast, not providing help led to negative evaluations of women but had little effect on the evaluation of men. That altruism is a requirement by the female gender-role finds also support by a meta-analysis of experiments on altruistic behavior: Facilitating intuitive, fast responses lead to more giving behavior by women, but not men (Rand et al., 2016).

Also in line with the altruistic female gender-role, Carli (1990) found that while for men, the speaking style made no difference, tentatively speaking women were more influential with men than confident, dominantly speaking women. A meta-analysis of studies regarding dominant behavior showed that dominant women, but not men, received decreased ratings in likeability and even hiring. The fact that women who violate prescriptive gender norms experience negative consequences is called the *backlash-effect*. It was coined by Rudman (1998), who found that when women engaged in self-promotion, competence ratings were increased, but simultaneous ratings of likeability were decreased. While this backlash occurs because of the violation of the communality norm, dominant women were still perceived as competent (Williams & Tiedens, 2016). Dealing out criticism puts women in a difficult position, as it violates prescriptive gender norms of being nice, understanding, and considerate.

## 3 Gender Differences in Leadership Behavior

A meta-analysis examining gender differences in actual leadership styles found that women lead more democratic or participative, while men more often adopt an autocratic leadership style (Eagly & Johnson, 1990). An optimal female leader is expected to be significantly more considerate of others than an optimal male leader (Russell et al., 1988), and being considerate of others is more important for being promoted for women than for men (Vinkenburt et al., 2011). In the specific field of academia, Clune (2009) examined college students expectations for a good male versus female professor. Positively evaluated male instructors were more often assertive, while positively evaluated female instructors were more often characterized as caring (see also *women-are-wonderful-effect* by Eagly & Mladinic, 1994). In line, *benevolent sexism* can predict gender inequality (Glick et al., 2000). It refers to a positive attitude towards women and stereotypes of women being inherently warmer, more caring, and morally superior to men, making them desirable as romantic partners and in need of a man’s protection. While the communal norm might lead to a positive attitude towards women, the characteristics that it ascribes to women do not qualify for high-status jobs, but for domestic tasks and care-taking (Eagly & Mladinic, 1994).

Referring to the before-mentioned backlash effect, Okimoto and Brescoll (2010) manipulated the gender and extent of power-seeking of political candidates. Female political candidates were voted for less when aiming for power, while power-seeking was no disadvantage for male candidates. Participants experienced feelings of moral outrage against the stereotype-violating woman, which mediated backlash (Brescoll et al., 2018). Based on this, we hypothesized: Workplace bullying accusations will lead to more moral outrage against a female leader than a male leader.

The backlash effect might serve as one explanation for the differences in leadership styles, as women often need to fear negative reactions when leading in similar styles as their male colleagues. Female leaders are evaluated more negatively than male leaders when they engage in autocratic and non-participative leadership styles (Eagly et al., 1992;

Rhee & Sigler, 2015). Moreover, company failure is blamed, or internally attributed, more often to female leaders with an autocratic leadership style than to any other combination of gender and leadership style (Lopez & Ensari, 2014). When female college instructors evaluated students negatively, they were rated as less competent than male instructors (Sinclair & Kunda, 2002). Moreover, Brescoll and Uhlmann (2008) found that angry women are conceded lower status than angry men. While men's anger emotions are attributed to external characteristics like the situation, women's anger is internally attributed. Negative reactions to women displaying anger are related to political conservatism and benevolent sexism (Salerno et al., 2018). In accordance, women who are depicted as managers, especially as successful ones, are evaluated more negatively regarding interpersonal attributes (e.g., interpersonal hostility and likeability) compared to women in general and, in some cases, male managers (Heilman et al., 1995; Heilman et al., 2004). This effect is substantially mitigated when there is information indicating that the woman also has communal characteristics (Heilman & Okimoto, 2007). The negative ratings for successful women go back to what was coined the implied communality deficit (Heilman et al., 2004). To be successful in a predominantly male field, it seems necessary to be agentic. If a woman is successful and therefore agentic, it is assumed that she is violating prescriptive norms to be communal. Therefore, successful female leaders are often perceived as cold and mean, because they are perceived as differing greatly from the picture of a typical woman. Female politicians who portray a dominant, male leadership style are often presented in the media as so-called *iron maidens* (Heilman et al., 2004). So, we hypothesized: The intention to punish a female leader accused of workplace bullying will be higher than for a male leader.

The circumstance that women are confronted with opposing expectations by gender norms and expectations for leadership positions has been called *double bind*. If women act communal, they are likely to be not considered for leadership positions. If they act agentic, they might not be liked (Rudman & Glick, 2001). This again has been shown to have very negative career outcomes, too (Heilman et al., 2004; Phelan et al., 2008). To avoid backlash and be successful, women need to walk a very fine line of conflicting norms (Rudman & Glick, 2001).

## 4 Female Leaders and Workplace Bullying Judgments

Workplace bullying can be defined as repeated, systematic harassing, excluding or impeding the work of the victim over a period of time, often with the perpetrator being in a position of power (Einarsen et al., 2011). Regarding research on workplace bullying and gender, there is evidence suggesting that men and women are victim of bullying at the workplace with a similar frequency (Einarsen & Skogstad, 1996), or women being bullied more often (Salin, 2015). Men are the perpetrators of workplace bullying more frequently (Einarsen & Skogstad, 1996), which is not surprising considering a majority of workplace bullying is conducted by supervisors, positions that are widely held by men (Rayner, 1997).

The relationship of female leaders and workplace bullying has not been studied explicitly before to our knowledge. A phenomenon that has received some attention is the so-called *queen bee syndrome* (Ellemers et al., 2004; Staines et al., 1974). A *queen bee* is defined as a successful woman who does not only not support, but perhaps even mistreats other subordinate women. It has been argued that women might develop this kind of behavior because they had to distance themselves from their gender identity and norms to be successful in male-dominated fields, and now act out the same treatment they received (Derks et al., 2011). However, there is convincing counter-evidence of female leaders promoting other women in leadership positions (Arvate et al., 2018). Also, the queen bee syndrome has been criticized for its implications. Mavin (2008) argues that it puts blame for the underrepresentation of women in management on successful female individuals, and reproduces stereotypes of successful women being mean, while there is no such label for the same behavior in male managers. Regardless of whether an effect can be found or not, it is striking how the picture of the queen bee fits the stereotypes of the communality deficit. Both concepts share the notion that female leaders are generally mean towards their subordinates and, therefore, very much differ from women in general.

The impact of gender on credibility and punishment has mostly been studied in the context of mock jury trials. For a defendant of a crime, it seems to be advantageous to be female (Mazzella & Feingold, 1994; Pozzulo et al., 2010). More feminine appearing defendants were evaluated more positively regarding likeability and credibility, which correlated with lower guilt ratings (Maeder & Dempsey, 2013). Outside of the criminal justice system there is also evidence for female speakers to be judged as more credible in daily-life (Robinson et al., 1998).

Those findings are not surprising, considering women are stereotyped as being communal. But how about female leaders? They are characterized as agentic and therefore lacking in communality. In line, we hypothesized, accusations of workplace bullying will be judged to be more accurate when they are made against a female leader than when made against a male leader.

While, to our knowledge, there is no research about women in leadership positions and credibility or punishment, it is known that human beings make judgements based on cognitive schemas and heuristics. Stereotypes can be considered cognitive schemas as well (Hamilton, 1979; Taylor et al., 1978). When stereotypes of the defendant and the crime match, this mediates judgment (Macrae & Shepherd, 1989). Smalarz et. al (2016) asked participants to compare the fingerprints found at a crime scene with those of a suspect. When the information about the suspect (sex or race) was stereotypical of the crime, participants judged the two fingerprints as matching significantly more often, even though they were not. Matching stereotypes also lead to a crime being more internally attributed to the defendant (Duncan, 1976). In Bodenhausen and Wyer's (1985) experiment, a person was accused of a job-related misdemeanor or crime that was either stereotypic for his ethnicity or not. When ethnicity and offense matched by a stereotype, participants judged the behavior as being more likely to reoccur and therefore even decided for a more severe punishment. Moreover, when we are presented with information that goes in line with stereotypes, we tend to process this information without great depth and without considering the credibility of the source (Macrae et al., 1992). While most stereotypes regarding crimes that have been examined relate to ethnicity, it seems plausible that the same goes for gender norms. The research finding by Grubb and Turner (2012) that female rape victims violating gender stereotypes have been given more blame than gender-norm confirming women supports this conclusion. In accordance, we hypothesized that a female leader is going to be judged as less credible than a male leader when accused of workplace bullying. Further, we assumed the hypothesis: A suspected victim will be judged to be more credible when accusing a female leader of workplace bullying than when accusing a male leader.

To test our hypotheses, we manipulated the gender in a scenario of a professor confronted with workplace bullying accusations and measured moral outrage, punishment, and credibility. Additionally, political orientation, benevolent sexism and gender-role were assessed (Rand et al., 2016; Salerno et al., 2018).

## 5 Method

### 5.1 Subjects

202 persons participated in the study, 141 women, 60 men and one person indicating their gender as diverse/nonbinary. In all analyses including gender of participants, the category diverse/nonbinary was excluded due to only one participant being in this category. The mean age was 35.04 years ( $SD = 14.23$ ). Of 202 participants, 81 were employed, 28 were self-employed and 73 were university students, 5 were civil servants, 5 in retirement, and 10 chose the answer option "other". Regarding compensation, 43 participants were second semester psychology students receiving credits for participating. The other participants were recruited mainly via social media and participated without receiving payment. Some participants not answering every question resulted in missing values only in the two items of the political orientation scale and some items of the BEM sex-role inventory. We chose to not fully exclude participants from our calculations based on this. Therefore, sample sizes vary between 184 and 202 subjects.

### 5.2 Design

The study was a one-factor between-subjects design, the factor being the gender of the accused leadership person (male/female). The participants were randomly assigned to one of the two conditions. The randomization was carried out automatically by the online survey platform.

### 5.3 Procedure & Material

An online study was conducted. Participants were to imagine themselves being in a commission of a German university, responsible for deciding over punishment for reported misbehavior of university employees. Their current case was an employee reporting their boss for workplace bullying. First, participants were presented with background information on the suspected victim (Mrs. Wagner), including a short summary of her career. The full version of the original material as well as a translated version can be found in the online supplementary material.<sup>1</sup>

Afterwards participants read the information on the suspected bully (Professor Dr. Schmidt), and also a short summary of his/her most recent career steps. Throughout all information given to the participants, the gender of the leader person and suspected bully was varied between the two conditions. Next, participants were presented with the statement of the suspected victim Mrs. Wagner. The statement includes a description of how she was treated by Professor Schmidt, accusing him/her of bullying. After that, participants were presented with the counter statement by the suspected bully, Professor Dr. Schmidt. The statement includes a description of Mrs. Wagner's performance and a direct answer to the bullying accusations from Professor Dr. Schmidt's perspective.

After reviewing all information, credibility of the suspected victim was assessed by asking "How credible do you consider Mrs. Wagner to be?". Credibility of the suspected bully was measured with the question "How credible do you consider Prof. Dr. Schmidt to be?". Both were measured on a 9-point scale from *not at all credible* (9) to *highly credible* (1). Then, participants were asked for a judgment whether they believed the bullying had happened ("Did, in your opinion, the supervisor Prof. Dr. Schmidt bully Mrs. Wagner?") on a 9-point scale from *no* (1) to *yes* (9) (*accuracy of bullying*). We created an additional measure for *accuracy of the accusations* by asking "To which extent do you agree to the following statements? The accusations are..." with the four items (Cronbach's  $\alpha = .68$ ) being "...justified", "...accurate", "...exaggerated" and "...unfounded" on a 9-point scale from *do not agree at all* (1) to *completely agree* (9). Following, to measure to which extent participants experience feelings of moral outrage against the suspected bully, we used the *moral outrage* items of Okimoto and Brescoll (2010). Drawing from research on moral emotions, the authors used seven items (Cronbach's  $\alpha = .92$ ) to assess the three primary "other-directed" moral emotions "contempt" (and "disdain"), "anger" (and "irritation" and "disapproval") and "disgust" (and "revulsion"). Participants were asked to indicate how much they felt these emotions towards the suspected bully Prof. Dr. Schmidt on a 7-point scale from *not at all* (1) to *very strongly* (7). Next, participants were asked about their intention to punish by removing the suspected bully Professor Dr. Schmidt from the leading position of the science center. We created an item asking participants to indicate their agreement to the statement "Prof. Dr. Schmidt should no longer be in a leading position at the science center." on a 9-point scale from *do not agree* (1) to *agree* (9).

Moreover, as a control variable, benevolent sexism attitudes of participants were measured using the German adaptation (Eckes & Six-Materna, 1999) of the benevolent sexism subscale of the *Ambivalent Sexism Scale* (Glick & Fiske, 1996). It includes 11 items (Cronbach's  $\alpha = .58$ ). Participants were to indicate their opinion about men and women and their relation towards another on a 6-point scale from *do not agree at all* (1) to *completely agree* (6). In addition, we assessed the individual gender-role of our participants by using the *BEM sex-role-inventory* (Bem, 1974), which serves to measure to which degree a person would describe themselves as being typically masculine or feminine. Masculinity and femininity can be considered two independent dimensions according to Bem (1974). A person is assessed to be either typically feminine or masculine when the difference score is high, or as having an androgynous profile when the difference score is low. We used a German revision of the scale (Troche & Rammsayer, 2011), with 15 items on each, the male (Cronbach's  $\alpha = .84$ ) and the female (Cronbach's  $\alpha = .90$ ), factor. Instead of dichotomous variables calculated by median splits, we used raw sex-role scores as control variables. Sex-role scores were calculated as average responses for each of both scales for each participant (Chung & Harmon, 1994; Conway et al., 1990). Participants were asked to describe themselves by indicating to which degree the items, e.g. "Acts as a leader" or "Affectionate" apply to them on a 7-point scale from *never, or almost never* (1) to *always* (7). Lastly, participants were requested to indicate their political orientation by stating the party they voted for at the last country wide election. The options were the major German parties in the parliament since the last election (*The Left, Alliance 90/the Greens, Social Democratic Party of Germany, Free Democratic Party, Christian Democratic Union of Germany, Alternative for Germany*), one collective category for all small parties under five percent of the votes, and an option to give no answer. Additionally, participants were asked to

<sup>1</sup> Material can be found in the OSF [https://osf.io/zt7vm/?view\\_only=0b98d43158594166bcd74968084719e2](https://osf.io/zt7vm/?view_only=0b98d43158594166bcd74968084719e2).



indicate where they would position themselves in the political spectrum on a 7-point scale from *left* (1) to *right* (7) and *liberal* (1) to *conservative* (7) (Cronbach's  $\alpha = .69$ ) (Janoff-Bulman et al., 2008).

## 6 Results

### 6.1 Testing of Hypotheses

Intercorrelations of dependent variables are presented in Table 1.<sup>2</sup>

**Table 1:** Bivariate correlations, means and standard deviations of individual measures and dependent variables.

Measures	1	2	3	4	5	6	7	8	9	10	11	M	SD
1. Victim credibility <sup>a</sup>	-											4.47	1.71
2. Bully credibility <sup>a</sup>	-.16*	-										4.22	1.66
3. Accuracy of bullying <sup>b</sup>	-.36**	.52**	-									4.24	2.14
4. Accuracy of accusations <sup>b</sup>	-.42**	.35**	.49**	-								5.15	1.27
5. Moral Outrage <sup>c</sup>	-.28**	.42**	.61**	.35**	-							2.76	1.34
6. Punishment intentions <sup>d</sup>	-.16*	.38**	.60**	.33**	.57**	-						2.93	2.06
7. Benevolent sexism <sup>e</sup>	-.10	-.05	.03	-.02	.15*	.04	-					2.82	0.81
8. Political orientation <sup>f</sup>	.07	-.12	.04	-.11	.12	-.04	.31**	-				3.26	1.13
9. Male sex-role factor <sup>g</sup>	-.04	-.13	-.01	-.06	-.01	-.03	.11	.13	-			4.32	0.87
10. Female sex-role factor <sup>g</sup>	.04	.00	.03	-.09	.04	.04	.09	.12	.23**	-		5.03	0.70
11. Age	-.07	.01	.03	-.10	.03	-.10	.23**	.30**	.04	.12	-	34.96	14.25
12. Gender	.04	.14*	.15*	.00	.11	.03	.14**	.02	-.16*	.24**	.04	-	-

*Note.* Sample size varied between  $N = 184$  and  $N = 202$ .

<sup>a</sup>Credibility of suspected victim and bully were assessed on a 9-point scale, with low numbers indicating high credibility. <sup>b</sup>Credibility of bullying and accusations were measured on 9-point scales, higher means indicating higher credibility. <sup>c</sup>Moral outrage ranged from 1 to 7, higher means indicating stronger feelings of moral outrage. <sup>d</sup>Punishment was assessed on 9-point scales, higher means indicating higher punishment. <sup>e</sup>Benevolent sexism of participants was measured on a 6-point scale, high numbers indicating strong sexism. <sup>f</sup>Political orientation was measured on a 7-point scale, 1 being left/liberal and 7 being right/conservative. <sup>g</sup>Male and female factor of the *BEM sex-role-inventory* were measured on a 7-point scale, high numbers indicating high masculinity/femininity.

\* $p < .05$ . \*\* $p < .01$ .

To test our hypotheses, we ran a multivariate ANOVA with gender of bully and gender of participants as independent variables and with moral outrage measures, punishment judgments, accusation judgments, and credibility judgments as dependent variables. The multivariate effect of gender of the suspected bully was not significant,  $F(6, 192) = 1.31$ ,  $p = .255$ . The multivariate effect of gender of participants was also not significant,  $F(6, 192) = 1.66$ ,  $p = .134$ . There was, moreover, no significant multivariate interaction between gender of participants and gender of suspected bully,  $F(6, 192) = .80$ ,  $p = .571$ . We ran separate univariate ANOVA's with gender of bully and gender of participants as independent variables on each dependent variable (see Table 2) to test our hypotheses in more detail.

*Moral outrage measure.* Inconsistent with our prediction, participants feeling of moral outrage against the suspected bully was not significantly effected by gender of the bully, female ( $M = 2.67$ ,  $SD = 1.31$ ) or male ( $M = 2.86$ ,  $SD = 1.36$ ),  $F(1,$

<sup>2</sup> Data and syntax file can be found in the OSF9 [https://osf.io/zt7vm/?view\\_only=0b98d43158594166bcd74968084719e2](https://osf.io/zt7vm/?view_only=0b98d43158594166bcd74968084719e2).

**Table 2:** Means and standard deviation of credibility, moral outrage and punishment judgments as a function of gender of the suspected bully and gender of the participants.

Measures	Female participants		Male participants	
	Female leader	Male leader	Female leader	Male leader
Victim credibility <sup>a</sup>	4.61 (1.52)	4.43 (1.81)	4.60 (1.83)	4.13 (1.80)
Bully credibility <sup>a</sup>	4.24 (1.55)	4.50 (1.77)	3.77 (1.70)	3.97 (1.56)
Accuracy of bullying <sup>b</sup>	4.25 (2.05)	4.63 (2.10)	3.63 (2.08)	3.90 (2.37)
Accuracy of accusations <sup>b</sup>	4.88 (1.15)	5.43 (1.19)	4.90 (1.51)	5.40 (1.33)
Moral Outrage <sup>c</sup>	2.71 (1.32)	2.99 (1.42)	2.56 (1.30)	2.54 (1.17)
Punishment <sup>d</sup>	3.08 (1.92)	2.84 (2.09)	2.63 (1.90)	3.03 (2.47)

Note. N = 201. Standard deviations in parentheses.

<sup>a</sup> Credibility of the suspected victim and the suspected bully was assessed on a 9-point scale, with low numbers indicating high credibility. <sup>b</sup> Both, credibility of the bullying and of the accusations were measured on 9-point scales, higher means indicating higher credibility. <sup>c</sup> Moral outrage ranges from 1 to 7, higher means indicating stronger feelings of moral outrage. <sup>d</sup> Punishment was assessed on 9-point scales, higher means indicating higher punishment.

199) = .44,  $p = .507$ . It was also not significantly effected by their own gender, male ( $M = 2.55$ ,  $SD = 1.23$ ) or female ( $M = 2.85$ ,  $SD = 1.37$ ),  $F(1, 199) = 2.13$ ,  $p = .146$ . There was no significant interaction between gender of suspected bully and gender of participants regarding the moral outrage measure,  $F(1, 199) = 0.54$ ,  $p = .464$  (see Table 2).

*Punishment judgments.* Against our hypothesis, participants did not intend to punish the suspected female bully ( $M = 2.95$ ,  $SD = 1.92$ ) significantly more than the suspected male bully ( $M = 2.90$ ,  $SD = 2.20$ ),  $F(1, 199) = .06$ ,  $p = .804$ . There was furthermore no significant difference between male ( $M = 2.83$ ,  $SD = 2.20$ ) and female ( $M = 2.96$ ,  $SD = 2.00$ ) participants punishment intentions,  $F(1, 199) = 0.17$ ,  $p = .683$ . Gender of participants and suspected bully did not significantly interact regarding punishment intentions,  $F(1, 199) = 1.02$ ,  $p = .315$  (see Table 2).

*Accusation judgments.* Participants did not significantly judge the bullying accusations to be more true in the female leader condition ( $M = 4.07$ ,  $SD = 2.07$ ) than in the male leader condition ( $M = 4.41$ ,  $SD = 2.20$ ),  $F(1, 199) = .96$ ,  $p = .328$ . Gender of participants had a significant effect on the truth judgments of bullying (accuracy of bullying),  $F(1, 199) = 4.25$ ,  $p = .041$ ,  $\eta^2 = .021$ . Women ( $M = 4.44$ ,  $SD = 2.08$ ) judged the bullying to be true significantly more than men ( $M = 3.77$ ,  $SD = 2.21$ ) did. There was no significant interaction of gender of suspected bully and gender of participants for accuracy of bullying,  $F(1, 199) = 0.27$ ,  $p = .869$  (see Table 2).

Moreover, the analysis yielded a significant effect of gender of suspected bully on accuracy of accusations,  $F(1, 199) = 7.45$ ,  $p = .007$ ,  $\eta^2 = .036$ . Contrary to our hypothesis, participants judged the accusations to be significantly more accurate in the male ( $M = 5.42$ ,  $SD = 1.23$ ) than in the female leader condition ( $M = 4.88$ ,  $SD = 1.26$ ). Female ( $M = 5.15$ ,  $SD = 1.20$ ) or male ( $M = 5.15$ ,  $SD = 1.43$ ) gender of participants did not have a significant effect on accuracy of accusation judgments,  $F(1, 199) = 0.00$ ,  $p = .989$ . There was no significant interaction of gender of suspected bully and gender of participants for accuracy of accusations,  $F(1, 199) = 0.02$ ,  $p = .893$  (see Table 2).

*Credibility judgments.* Regarding credibility of the suspected victim, no significant effect of male ( $M = 4.34$ ,  $SD = 1.80$ ) or female ( $M = 4.60$ ,  $SD = 1.61$ ) gender of bully was found,  $F(1, 199) = 1.49$ ,  $p = .224$ . There was no significant difference between credibility judgments for the suspected victim of male ( $M = 4.37$ ,  $SD = 1.81$ ) and female ( $M = 4.52$ ,  $SD = 1.66$ ) participants,  $F(1, 199) = 0.33$ ,  $p = .569$ . We found no significant interaction between gender of suspected bully and gender of participants on credibility of suspected victim,  $F(1, 199) = .30$ ,  $p = .584$ . (see Table 2).

Against our predictions, we found no significant effect of female ( $M = 4.10$ ,  $SD = 1.60$ ) or male ( $M = 4.34$ ,  $SD = 1.72$ ) gender of suspected bully on credibility of suspected bully,  $F(1, 199) = .82$ ,  $p = .367$ . There was however a significant effect of gender of participants on credibility judgments of suspected bully,  $F(1, 199) = 3.90$ ,  $p = .05$ ,  $\eta^2 = .019$ . Male participants rated the suspected bully as more credible ( $M = 3.87$ ,  $SD = 1.62$ ) than female participants ( $M = 4.37$ ,  $SD =$

1.66) in both conditions. There was no significant interaction between gender of the bully and gender of suspected participants for credibility of suspected bully,  $F(1, 199) = .01, p = .906$  (see Table 2).

## 6.2 Individual Differences: Benevolent Sexism, Gender-role and Political Orientation

Pearson's correlations, means, and standard deviations of individual measures of participants are presented in Table 1. Male participants were found to hold more sexist attitudes than female participants, which is consistent with the findings of Glick and Fiske (1996). Moreover, in line with previous work (Glick et al., 2002), older participants indicated more sexist attitudes than younger participants. Older participants were also assessed to be more politically conservative, as expected based on earlier studies (Salerno et al., 2018; Salerno & Phalen, 2019). More politically conservative participants were found to score significantly higher on the benevolent sexism measure. Previous studies have also found a positive correlation between benevolent sexism and political conservatism (Salerno et al., 2018; Salerno & Phalen, 2019). There was further a significant correlation between the average raw scores of the BEM sex-role factors and the respective gender. As expected, both subscales of the *BEM sex-role-inventory* correlated positively, as masculinity and femininity can be considered two independent dimensions, consistent with Bem's (1974) findings.

When correlating the dependent measures with individual difference measures of participants, a positive significant correlation between sexism and moral outrage was found,  $r(198) = .15, p = .039$ . Higher individual sexism was correlated with more moral outrage against the suspected bully, no matter the bully's gender. Benevolent sexism includes a protective attitude towards women, and the victim was female in both conditions. Benevolent sexism was not significantly correlated with credibility, accuracy, moral outrage or punishment intention measures (see Table 1). For political orientation, there was also no significant correlation with the dependent measures found, and neither BEM's male nor female factor correlated with the dependent measures (see also Table 1). In line with the results of hypotheses testing, gender of participants was significantly positively correlated with credibility of suspected bully,  $r(199) = .14, p = .050$ . Also a significant positive correlation of gender with accuracy of bullying was found,  $r(199) = .15, p = .041$ . Women judged the suspected bully as less credible and the bullying as more true than men did. Age did not significantly correlate with the dependent measures (see also Table 1).

To further test possible moderation effects of individual difference measures on the effect of gender of suspected bully on dependent variables, we ran several linear regression models. We only found one significant moderation effect. A significant regression was found for the model with accuracy of accusations as a dependent variable and gender of suspected bully, male sex-role score and their interaction as predictors,  $R^2 = .068, F(3, 193) = 4.70, p = .003$ . Further a significant main effect of gender of suspected bully on accuracy of accusations was shown,  $b = -2.42, t(3, 196) = -2.65, p = .009$ . Moreover, we found a significant moderation effect of gender of suspected bully and male sex-role score regarding accuracy of the accusations,  $b = .44, t(196) = 2.11, p = .036$ . The higher participants scored on male sex-role factor, the less accurate they found the accusations against the male leader to be ( $b = -.27, t(196) = -1.96, p = .051$ ). For the female leader, a positive but not significant slope of male sex-role and accuracy of accusations was found ( $b = .167, t(196) = 1.07, p = .282$ ). The lower participants scored on male sex-role factor, the more accurate they found the accusations against the male leader, and the less accurate they found the accusations against the female leader to be. Further, the model predicting accuracy of accusations with gender of suspected bully, benevolent sexism and their interaction was significant,  $R^2 = .051, F(3, 196) = 3.50, p = .017$ . There was a significant main effect of gender of suspected bully on accuracy of accusations,  $b = -1.49, t(196) = -2.33, p = .021$ . No other moderator effect was found. All results of multiple regression models can be found in Appendix A.

## 7 Discussion

The present study examined whether a female leader would be evaluated less favorably compared to a male leader when accused of bullying a female victim at the workplace. The results did not go into the expected direction but offer some interesting secondary findings that may inform the relationships hypothesized.



Participants did not indicate higher feelings of moral outrage against and punishment intentions for a female leader compared to a male leader accused of workplace bullying. This can be considered surprising, as past research would have suggested differently. Women were found to experience backlash effects when violating prescriptive gender norms (Brescoll et al., 2018; Okimoto & Brescoll, 2010; Rudman, 1998), such as being communal (Burgess & Borgida, 1999; Eagly, 1987). Furthermore, a female leader accused of workplace bullying was not judged to be significantly less credible (compared to a male leader). In addition, the employee accusing a female leader of bullying was not judged to be more credible compared to the employee accusing a male leader. Referring to recent research, female leaders were found to be stereotyped as lacking in communality (Heilman et al., 1995; Heilman et al., 2004; Heilman & Okimoto, 2007), which we hypothesized to fit with workplace bullying accusations and therefore lead to more negative judgements (Bodenhausen & Wyer, 1985; Macrae et al., 1992; Smalarz et al., 2016). In previous work, judgments were affected by individual difference measures like benevolent sexism and political orientation (Salerno et al., 2018; Salerno & Phalen, 2019). We did not find benevolent sexism or political orientation to affect judgments. Therefore, our study did not provide evidence for a discrimination against female leaders accused of workplace bullying.

Contrary to our predictions, we found that the workplace bullying accusations against a male leader were significantly judged to be more accurate than accusations against a female leader. As is often the case in stereotype research, one attempt at explanation would be to argue that participants were influenced by social desirability. This would be especially problematic given a within subject design with both male and female target person, in which participants would experience the social desire to show no discrimination against the woman compared to the man. Due to our between-subjects factorial design this seems more unlikely, as the purpose of the study was probably less obvious to the participants, but it is not impossible that social desirability had an influence on our results. Another explanation might be a different stereotype. In our experiment, the victim in the scenario was female in both conditions and the background information implied that the suspected victim just graduated college, while the suspected bully, being a professor and holding a leading position, had to be much older. In the male leader condition, the scenario therefore was a young woman accusing a substantially older boss of mistreating her at work. This constellation might have reminded participants of a stereotype of an older male boss as a perpetrator, which may have been a more easily accessible stereotype in the context of workplace bullying than the one of the mean and cold female boss. Sexual abuse and bullying fall in the same category of interpersonal mistreatment, but sexual abuse has been much more prevalent in the media in the last years (Lim & Cortina, 2005). One example is the *#MeToo* movement, in which women started to publicly speak up about sexual abuse by powerful men. Moreover, our finding is also in line with research mock-jury trials. A meta-analysis found, it is disadvantageous for a defendant to be male, and it is also disadvantageous for him if the victim of the crime is female (Mazzella & Feingold, 1994). Based on this research, our male leader condition was the most likely one for the defendant to be judged guilty. Presumptively, participants perceived the crime of mistreating a young female victim at the workplace and the defendant being an older male boss as matching in stereotype (Bodenhausen & Wyer, 1985; Macrae et al., 1992; Smalarz et al., 2016), and therefore judged the accusations against the male leader to be more accurate. The results might have been more in line with our hypothesis if we had controlled for or manipulated the seniority of the accused bully differently. In line with matching stereotypes, a manipulation check regarding the perception of the bully being agentic versus communal seems meaningful. It could be that the male leader was still perceived as more agentic and less communal than the female leader and therefore the bullying accusations against him were judged as more accurate.

Moreover, female participants judged the bullying accusations to be more credible than male participants no matter the gender of the bully. Male participants generally found the leader, male or female, to be more credible than female participants. So, female participants were more warily in judging against the victim. Perhaps, men sympathized more with the bully, while women felt more for the victim. Future studies should further examine this. Nevertheless, the finding is in line with research on mock-jury trials, in which men found the defendants to be more credible than women did, while women judged the victims to be more credible than men did (Pozzulo et al., 2010). As there were only two conditions in our study, the suspected victim was always female. Research on secondary gender-based mistreatment (Miner & Eischeid, 2012) found that participants indicated more negative feelings when they observed someone with the same gender being mistreated, an effect that might also have influenced our finding. Therefore, we cannot derive from the results that in general women show more and men less empathy towards victims.

We also found that participants with higher male sex-role scores judged the accusations against the male leader to be less accurate, and the accusations against the female leader to be more accurate. It can be wondered why male

participants seemed to sympathize more with the male leader, while no such effect was found for female participants and the female leader. Participants with a higher male sex-role score might have been more aware of the stereotype of a man with power as a typical perpetrator, and therefore might have anticipated more that the accusations could be false, or his behavior could have been misinterpreted based on this stereotype. It is also possible that judging the accusations against the male leader to be less accurate serves in protecting the high male sex-role scores participants self-worth.

## 8 Limitations and Future Research

While the results are surprising, our study has some limitations. At first, we want to highlight that we tested our hypotheses in only one study with just one specific hypothetical scenario. In this scenario, one limitation of our experiment was that the gender of the victim was female in both conditions, as we decided to focus our limited resources on varying the gender of the suspected bully. Considering our findings, this might have influenced our results. Additionally, our sample was predominantly female. One can speculate that primarily male or male gender-typed groups might show more negative judgments towards female leaders accused of workplace bullying. This is especially relevant, as gender and sex-role of participants both were shown to affect judgments. For future research, it would be interesting to implement a two-factorial-design, varying gender of both victim and bully, with a sample more balanced regarding gender. Results of a *female bully-male victim* and *male bully-male victim* condition might lead to new insights.

We did not include a manipulation check to test whether participants perceived the female leader to be more agentic than the male one. As earlier mentioned, this could be a possible reason why the accusations against the male leader were judged to be more accurate. Based on earlier research regarding the communality deficit, we presupposed, that female leaders could be perceived as having a communality deficit in comparison to male leaders. In line with our results, it would also be plausible that female leaders are perceived as less communal than females in non-leading roles. This assumption is according to the finding by Heilman et al. (1995; 2004) who showed that interpersonal attributes of female managers (especially successful ones) were evaluated negatively compared to women in general. We suggest for future research to test whether female leaders are more easily accused of bullying than female non-leaders.

An alternative hypothesis not examined by our study could be, that female leaders are being accused of workplace bullying more often due to a different male-female perception, namely, that women are not as aggressive as men. One could argue that the accusation could be easier brought forward against females as not as much resistance is to be anticipated. In our study design, participants were asked to judge accusations already made. Future studies should examine this alternative mechanism with a design testing whether more accusations are brought forward against a female leader, and whether participants have different expectations regarding the leader's reaction to the accusation depending on gender. As the last point, research on the intersection of gender and race is another important aspect future studies should consider (Livingston et al., 2012; Rosette et al., 2016; Rosette & Livingston, 2012).

## 9 Conclusion

To our knowledge, this was the first experiment examining female leadership and workplace bullying. The results can be summarized as follows: First, male leaders were more often judged to be workplace bullies compared to female leaders. Interestingly, participants with higher male sex-role scores judged the accusations against the male leader to be less accurate, and the accusations against the female leader to be more accurate. Further, female participants judged the victim as more credible compared to male participants, independent from the gender of the bully, substantially more research on the topic is needed.

**Funding:** No funding was received to assist with the preparation of this manuscript.

**Conflicts of interests/competing interests:** The authors have no relevant financial or non-financial interests to disclose.

**Informed consent:** Informed consent was obtained from all individual participants included in the study.

**Code availability/Availability of data and material:** Data and syntax file are uploaded in a OSF [https://osf.io/zt7vm/?view\\_only=0b98d43158594166bcd74968084719e2](https://osf.io/zt7vm/?view_only=0b98d43158594166bcd74968084719e2).

**Author's contribution statement:** The article is based primarily on the bachelor thesis conducted by the first author under supervision and mentoring by the second author. Research idea and theoretical hypotheses: Main contribution by the first author and additional contribution by the second author for the first version of the paper. In the revised final version: main contribution of the third author and additional contribution of the first and second author. Development of material and data collection: Main contribution by the first author and additional contribution by the second author. Data analysis: First and second author made equal contribution for the first version of the paper. In the revised final version main contribution of the third author and additional contribution of the first and second author. Discussion: Main contribution by the first author, additional contribution by the second author for the first version of the paper. In the revised final version main contribution of the third author and additional contribution of the first and second author.

The study was preregistered at AsPredicted. The preregistration can be found with the following link: <https://aspredicted.org/blind.php?x=fv2wm6>

## References

- Arvate, P. R., Galilea, G. W., & Todescat, I. (2018). The queen bee: A myth? The effect of top-level female leadership on subordinate females. *The Leadership Quarterly*, 29(5), 533–548. <https://doi.org/10.1016/j.leaqua.2018.03.002>
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162.
- Bodenhausen, G. V., & Wyer, R. S. (1985). Effects of stereotypes in decision making and information-processing strategies. *Journal of Personality and Social Psychology*, 48(2), 267–282. <https://doi.org/10.1037/0022-3514.48.2.267>
- Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour. *Economics Letters*, 170, 19–23. <https://doi.org/10.1016/j.econlet.2018.05.022>
- Brescoll, V. L., Okimoto, T. G., & Vial, A. C. (2018). You've come a long way... maybe: How moral emotions trigger backlash against women leaders. *Journal of Social Issues*, 74(1), 144–164. <https://doi.org/10.1111/josi.12261>
- Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, 19(3), 268–275. <https://doi.org/10.1111/j.1467-9280.2008.02079.x>
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692. <https://doi.org/10.1037/1076-8971.5.3.665>
- Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology*, 59(5), 941. <https://doi.org/10.1037/0022-3514.59.5.941>
- Chung, Y. B., & Harmon, L. W. (1994). The career interests and aspirations of gay men: How sex-role orientation is related. *Journal of Vocational Behavior*, 45(2), 223–239. <https://doi.org/10.1006/jvbe.1994.1033>
- Clune, K. F. (2009). *Students' perceptions of instructor credibility: Effects of instructor sex, gender role, and communication style* [Doctoral dissertation, University of Kansas]. KU ScholarWorks. <https://kuscholarworks.ku.edu/handle/1808/5973>
- Conway, M., Giannopoulos, C., & Stiefenhofer, K. (1990). Response styles to sadness are related to sex and sex-role orientation. *Sex Roles*, 22(9-10), 579–587. <https://doi.org/10.1007/BF00288236>
- Derks, B., Ellemers, N., Van Laar, C., & De Groot, K. (2011). Do sexist organizational cultures create the queen bee? *The British Journal of Social Psychology*, 50(3), 519–535. <https://doi.org/10.1348/014466610X525280>
- Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, 34(4), 590. <https://doi.org/10.1037/0022-3514.34.4.590>
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Psychology Press.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 108(2), 233. <https://doi.org/10.1037/0033-2909.108.2.233>
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111(1), 3. <https://doi.org/10.1037/0033-2909.111.1.3>
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15(4), 543–558. <https://doi.org/10.1177/0146167289154008>
- Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology*, 5(1), 1–35. <https://doi.org/10.1080/14792779543000002>

- Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus. *Zeitschrift Für Sozialpsychologie*, 30(4), 211–228. <https://doi.org/10.1024//0044-3514.30.4.211>
- Einarsen, S., Hoel, H., Zapf, D., & Cooper, C. L. (2011). The concept of bullying and harassment at work: The European tradition. In S. Einarsen, H. Hoel, D. Zapf, & C. L. Cooper (Eds.), *Bullying and harassment in the workplace: Developments in theory, research, and practice*. (Vol. 2, pp. 3–40). CRC Press.
- Einarsen, S., & Skogstad, A. (1996). Bullying at work: Epidemiological findings in public and private organizations. *European Journal of Work and Organizational Psychology*, 5(2), 185–201. <https://doi.org/10.1080/13594329608414854>
- Ellemers, N., Van den Heuvel, H., De Gilder, D., Maass, A., & Bonvini, A. (2004). The underrepresentation of women in science: Differential commitment or the queen bee syndrome?. *British Journal of Social Psychology*, 43(3), 315–338. <https://doi.org/10.1348/0144666042037999>
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Glick, P., Lameiras, M., & Castro, Y. R. (2002). Education and catholic religiosity as predictors of hostile and benevolent sexism toward women and men. *Sex Roles*, 47(9-10), 433–441. <https://doi.org/10.1023/A:1021696209949>
- Grubb, A., & Turner, E. (2012). Attribution of blame in rape cases: A review of the impact of rape myth acceptance, gender role conformity and substance use on victim blaming. *Aggression and violent behavior*, 17(5), 443-452.
- Hamilton, D. L. (1979). A cognitive-attribitional analysis of stereotyping. *Advances in Experimental Social Psychology*, 12, 53–84. [https://doi.org/10.1016/S0065-2601\(08\)60259-2](https://doi.org/10.1016/S0065-2601(08)60259-2)
- Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality*, 10(4), 237.
- Heilman, M. E., & Chen, J. J. (2005). Same behavior, different consequences: reactions to men's and women's altruistic citizenship behavior. *The Journal of Applied Psychology*, 90(3), 431–441. <https://doi.org/10.1037/0021-9010.90.3.431>
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *The Journal of Applied Psychology*, 92(1), 81–92. <https://doi.org/10.1037/0021-9010.92.1.81>
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *The Journal of Applied Psychology*, 89(3), 416–427. <https://doi.org/10.1037/0021-9010.89.3.416>
- Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology*, 10, 11.
- Innocenti, A., & Paziienza, M. G. (2006). Altruism and gender in the trust game. *Labsi Working Paper*, 5. <https://doi.org/10.2139/ssrn.884378>
- Janoff-Bulman, R., Sheikh, S., & Baldacci, K. G. (2008). Mapping moral motives: Approach, avoidance, and political orientation. *Journal of Experimental Social Psychology*, 44(4), 1091–1099. <https://doi.org/10.1016/j.jesp.2007.11.003>
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, 9, 1086.
- Lim, S., & Cortina, L. M. (2005). Interpersonal mistreatment in the workplace: The interface and impact of general incivility and sexual harassment. *The Journal of Applied Psychology*, 90(3), 483–496. <https://doi.org/10.1037/0021-9010.90.3.483>
- Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an agentic Black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychological Science*, 23(4), 354–358.
- Lopez, E. S., & Ensari, N. (2014). The effects of leadership style, organizational outcome, and gender on attributional bias toward leaders. *Journal of Leadership Studies*, 8(2), 19–37. <https://doi.org/10.1002/jls.21326>
- Macrae, C. N., & Shepherd, J. W. (1989). Do criminal stereotypes mediate juridic judgements?. *British Journal of Social Psychology*, 28(2), 189–191. <https://doi.org/10.1111/j.2044-8309.1989.tb00860.x>
- Macrae, C. N., Shepherd, J. W., & Milne, A. B. (1992). The effects of source credibility on the dilution of stereotype-based judgments. *Personality and Social Psychology Bulletin*, 18(6), 765–775. <https://doi.org/10.1177/0146167292186013>
- Maeder, E. M., & Dempsey, J. L. (2013). A likely story? The influence of type of alibi and defendant gender on juror decision-making. *Psychiatry, Psychology and Law*, 20(4), 543–552. <https://doi.org/10.1080/13218719.2012.727066>
- Mavin, S. (2008). Queen bees, wannabees and afraid to bees: No more 'best enemies' for women in management? *British Journal of Management*, 19(1), 75-84. <https://doi.org/10.1111/j.1467-8551.2008.00573.x>
- Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis 1. *Journal of Applied Social Psychology*, 24(15), 1315–1338. <https://doi.org/10.1111/j.1559-1816.1994.tb01552.x>
- Miner, K. N., & Eischeid, A. (2012). Observing incivility toward coworkers and negative emotions: Do gender of the target and observer matter?. *Sex Roles*, 66(7-8), 492–505. <https://doi.org/10.1007/s11199-011-0108-0>
- Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality & Social Psychology Bulletin*, 36(7), 923–936. <https://doi.org/10.1177/0146167210371949>
- Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, 32(4), 406–413. <https://doi.org/10.1111/j.1471-6402.2008.00454.x>
- Pozzulo, J. D., Dempsey, J., Maeder, E., & Allen, L. (2010). The effects of victim gender, defendant gender, and defendant age on juror decision making. *Criminal Justice and Behavior*, 37(1), 47–63. <https://doi.org/10.1177/0093854809344173>



- 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(4), 269–281. <https://doi.org/10.1111/1471-6402.t01-1-00066>
- Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology. General*, 145(4), 389–396. <https://doi.org/10.1037/xge0000154>
- Rayner, C. (1997). The incidence of workplace bullying. *Journal of Community & Applied Social Psychology*, 7(3), 199–208. [https://doi.org/10.1002/\(SICI\)1099-1298\(199706\)7:3<199::AID-CASP418>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1099-1298(199706)7:3<199::AID-CASP418>3.0.CO;2-H)
- Rhee, K. S., & Sigler, T. H. (2015). Untangling the relationship between gender and leadership. *Gender in Management.*, 30(2), 109–134. <https://doi.org/10.1108/GM-09-2013-0114>
- Robinson, K. A., Obler, L. K., Boone, R. T., Shane, H., Adamjee, R., & Anderson, J. (1998). Gender and truthfulness in daily life situations. *Sex Roles*, 38(9-10), 821–831. <https://doi.org/10.1023/A:1018877215432>
- Rosette, A. S., Koval, C. Z., Ma, A., & Livingston, R. (2016). Race matters for women leaders: Intersectional effects on agentic deficiencies and penalties. *The Leadership Quarterly*, 27(3), 429–445.
- Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects of organizational performance on leaders with single versus dual-subordinate identities. *Journal of Experimental Social Psychology*, 48(5), 1162–1167.
- Rubner, J. (2020, January 30). Die Angeklagten. *Die Zeit*, 6/2020. <https://www.zeit.de/2020/06/mobbing-professorin-vorwurf-beleidigungen-gerichtsprozess>
- 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(3), 629. <https://doi.org/10.1037/0022-3514.74.3.629>
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4), 743–762. <https://doi.org/10.1111/0022-4537.00239>
- Russell, J. E., Rush, M. C., & Herd, A. M. (1988). An exploration of women's expectations of effective male and female leadership. *Sex Roles*, 18(5-6), 279–287. <https://doi.org/10.1007/BF00288290>
- Salerno, J. M., & Phalen, H. J. (2019). Traditional gender roles and backlash against female attorneys expressing anger in court. *Journal of Empirical Legal Studies*, 16(4), 909–932. <https://doi.org/10.1111/jels.12238>
- Salerno, J. M., Phalen, H. J., Reyes, R. N., & Schweitzer, N. J. (2018). Closing with emotion: The differential impact of male versus female attorneys expressing anger in court. *Law and Human Behavior*, 42(4), 385. <https://doi.org/10.1037/lhb0000292>
- Salin, D. (2015). Risk factors of workplace bullying for men and women: The role of the psychosocial and physical work environment. *Scandinavian Journal of Psychology*, 56(1), 69–77. <https://doi.org/10.1111/sjop.12169>
- Sinclair, L., & Kunda, Z. (2002). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26(11), 1329–1342. <https://doi.org/10.1177/0146167200263002>
- Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. (2016). The perfect match: Do criminal stereotypes bias forensic evidence analysis? *Law and Human Behavior*, 40(4), 420–429. <https://doi.org/10.1037/lhb0000190>
- Staines, G., Tavis, C., & Jayaratne, T. E. (1974). The queen bee syndrome. *Psychology Today*, 7, 55–60. <https://doi.org/10.1037/e400562009-003>
- Taylor, S. E., Fiske, S. T., Etoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36(7), 778. <https://doi.org/10.1037/0022-3514.36.7.778>
- Troche, S. J., & Rammsayer, T. H. (2011). Eine Revision des deutschsprachigen Bem Sex-Role Inventory. *Klinische Diagnostik Und Evaluation*, 4, 262–283.
- Vinkenburg, C. J., Van Engen, M. L., Eagly, A. H., & Johannesen-Schmidt, M. C. (2011). An exploration of stereotypical beliefs about leadership styles: Is transformational leadership a route to women's promotion?. *The Leadership Quarterly*, 22(1), 10–21. <https://doi.org/10.1016/j.leaqua.2010.12.003>
- Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study*, Rev. Sage Publications, Inc.
- Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, 142(2), 165–197. <https://doi.org/10.1037/bul0000039>



## Appendix A

Multiple linear regression analysis of individual difference measures, gender of suspected bully and their interaction to predict dependent variables

Variables	$R^2$	$F$	$p$	$b$	$t$	$p$
Dependent variable: <i>Victim credibility</i>	.02	1.08	.359			
Gender of bully				-.05	-.06	.951
Benevolent sexism				-.28	-1.31	.193
Gender of bully x benevolent sexism				.11	.37	.714
Dependent variable: <i>Victim credibility</i>	.02	1.28	.281			
Gender of bully				1.35	1.72	.087
Political orientation				.24	1.65	.100
Gender of bully x political orientation				-.35	-1.53	.128
Dependent variable: <i>Victim credibility</i>	.01	.42	.741			
Gender of bully				.49	.39	.700
Male sex-role factor				-.05	-.25	.801
Gender of bully x male sex-role factor				-.06	-.20	.844
Dependent variable: <i>Victim credibility</i>	.01	.73	.535			
Gender of bully				1.37	.77	.441
Female sex-role factor				.21	.87	.388
Gender of bully x female sex-role factor				-.22	-.61	.540
Dependent variable: <i>Bully credibility</i>	.01	.49	.690			
Gender of bully				-.08	-.09	.931
Benevolent sexism				-.07	-.35	.727
Gender of bully x benevolent sexism				-.05	-.18	.858
Dependent variable: <i>Bully credibility</i>	.03	1.76	.156			
Gender of bully				-1.18	-1.57	.118
Political orientation				-.30	-2.18	.030*
Gender of bully x political orientation				.31	1.42	.159
Dependent variable: <i>Bully credibility</i>	.02	1.56	.202			
Gender of bully				-.13	-.10	.917
Male sex-role factor				-.23	-1.27	.205
Gender of bully x male sex-role factor				-.03	-.12	.907
Dependent variable: <i>Bully credibility</i>	.01	.39	.759			
Gender of bully				-.64	-.37	.714
Female sex-role				-.03	-.14	.886
Gender of bully x female sex-role				.08	.22	.824
Dependent variable: <i>Accuracy of bullying</i>	.009	.62	.600			
Gender of bully				-1.20	-1.08	.281

Variables	$R^2$	$F$	$p$	$b$	$t$	$p$
Benevolent sexism				-.08	-.03	.777
Gender of bully x benevolent sexism				.31	.83	.408
Dependent variable: <i>Accuracy of bullying</i>	.04	2.22	.087			
Gender of bully				-2.40	-2.52	.013*
Political orientation				-.19	-1.08	.283
Gender of bully X political orientation				.66	2.39	.018*
Dependent variable: <i>Accuracy of bullying</i>	.02	1.45	.229			
Gender of bully				-3.01	-1.97	.050*
Male sex-role factor				-.29	-1.22	.223
Gender of bully x male sex-role factor				.64	1.80	.073
Dependent variable: <i>Accuracy of bullying</i>	.02	1.42	.238			
Gender of bully				-4.20	-1.89	.060
Female sex-role factor				-.26	-.88	.380
Gender of bully x female sex-role factor				.78	1.78	.076
Dependent variable: <i>Accuracy of accusations</i>	.05	3.50	.017*			
Gender of bully				-1.49	-2.33	.021*
Benevolent sexism				-.18	-1.17	.244
Gender of bully x benevolent sexism				.354	1.63	.106
Dependent variable: <i>Accuracy of accusations</i>	.05	3.26	.023*			
Gender of bully				-.44	-.81	.417
Political orientation				-.11	-1.1	.245
Gender of bully x political orientation				-.009	-.060	.953
Dependent variable: <i>Accuracy of accusations</i>	.07	4.70	.003*			
Gender of bully				-2.42	-2.65	.009*
Male sex-role factor				-.27	-1.96	.051
Gender of bully x male sex-role factor				.44	2.11	.036*
Dependent variable: <i>Accuracy of accusations</i>	.05	3.39	.019*			
Gender of bully				-.92	-.71	.480
Female sex-role factor				-.19	-1.09	.275
Gender of bully x female sex-role factor				.08	.76	.758
Dependent Variable: <i>Moral outrage</i>	.03	2.01	.113			
Gender of bully				-.69	-1.00	.318
Benevolent sexism				.17	1.00	.315
Gender of bully x benevolent sexism				.17	.73	.468
Dependent Variable: <i>Moral outrage</i>	.02	1.36	.256			
Gender of bully				-.62	-1.02	.309
Political orientation				.09	.79	.431
Gender of bully x political orientation				.15	.82	.413
Dependent Variable: <i>Moral outrage</i>	.01	.32	.810			

Variables	<i>R</i> <sup>2</sup>	<i>F</i>	<i>p</i>	<i>b</i>	<i>t</i>	<i>p</i>
Gender of bully				.02	.02	.986
Male sex-role factor				.01	.08	.933
Gender of bully x male sex-role factor				-.05	-.21	.836
Dependent Variable: <i>Moral outrage</i>	.01	.31	.821			
Gender of bully				-.13	-.10	.925
Female sex-role factor				.08	.44	.662
Gender of bully x female sex-role factor				-.00	-.01	.993
Dependent Variable: <i>Punishment intentions</i>	.00	.13	.941			
Gender of bully				-.00	-.00	.997
Benevolent sexism				.11	.41	.682
Gender of bully x benevolent sexism				.01	.03	.974
Dependent Variable: <i>Punishment intentions</i>	.01	.43	.734			
Gender of bully				-.87	-.90	.367
Political orientation				-.18	-1.04	.299
Gender of bully x political orientation				.28	.99	.324
Dependent Variable: <i>Punishment intentions</i>	.03	1.64	.182			
Gender of bully				-3.12	-2.07	.040*
Male sex-role factor				-.40	-1.74	.083
Gender of bully x male sex-role factor				.74	2.16	.032*
Dependent Variable: <i>Punishment intentions</i>	.00	.23	.875			
Gender of bully				-.86	-.40	.687
Female sex-role factor				.04	.13	.897
Gender of bully x female sex-role factor				.19	.45	.650

Note. Sample size varied between N = 185 and N = 202.

\**p* < .05. \*\**p* < .01.