

FEMININITY IS FAVORABLE: SEXUALLY DIMORPHIC FACIAL FEATURES AFFECT ASSESSMENTS OF WHITE WOMEN'S LEADERSHIP ABILITIES

Hayley A. Liebenow
University of Toronto

Kathryn L. Boucher
University of Indianapolis

Brittany S. Cassidy
University of North Carolina at Greensboro

Women remain underrepresented in leadership positions traditionally held by men. Research on role congruity and backlash has shown that aspiring women versus men leaders are more negatively evaluated when they enact agentic behaviors. We examined whether sexually dimorphic facial features, which are associated with agentic and communal trait impressions, constitute a nonverbal barrier to White women's leadership in a college setting. Manipulated masculinized versus feminized facial features elicited, respectively, higher dominance and lower warmth impressions. Aspiring women leaders with masculinized versus feminized facial features received less favorable evaluations for several leadership roles, whereas men's evaluations were unaffected by varying features. Contrasting past work, aspiring women leaders were overall more favorably evaluated. This difference related to beliefs that college-aged White women versus men are more competent, responsible, and warm. These findings provide novel evidence that feminized facial features offer a unique advantage to aspiring college-aged White women leaders.

Keywords: face perception, gender stereotypes, sexual dimorphism, role congruity

Women are underrepresented in leadership in workplace domains ranging from science, technology, engineering, and mathematics (STEM) (Kahn & Ginther, 2017) to politics (Baskaran & Hessami, 2018) to academia (Llorens et al., 2021).

Supplementary materials are available online.

Address correspondence to Hayley A. Liebenow, Department of Psychology, University of Toronto, 100 St. George St., Toronto, ON M5S 3G3, Canada. E-mail: haliebenow@uncg.edu

This underrepresentation undermines women's sense of belonging (Baskaran & Hessami, 2018; Broockman, 2014; Creamer, 2012; Piatek-Jimenez et al., 2018), perpetuating gender disparity in positions traditionally held by men (Gadassi & Gati, 2009; Glick et al., 1995). Much research has focused on identifying causes of this underrepresentation. Influential in this research is role congruity theory (Eagly & Karau, 2002), whereby the perceived incongruity between the communal female gender role and the agentic characteristics of leadership causes evaluations of leadership-related behaviors to be negatively evaluated when women (versus men) enact them (Banchefsky et al., 2016; Gervais & Hillard, 2011). Paralleling such patterns are the social and economic penalties reflective of *backlash* (Rudman & Phelan, 2008), in that people like women less when they display more (versus fewer) agentic behaviors supporting their competency. Most work from these perspectives involves reading resumes or vignettes (e.g., Phelan et al., 2008; Rudman & Phelan, 2008). People, however, also make agentic and communal inferences from facial features (e.g., Sutherland et al., 2015). The current work examined the underexplored possibility that facial features elicit outcomes for women consistent with patterns expected from these perspectives.

Underlying tenets of role congruity theory (Eagly & Karau, 2002) and backlash discrimination research (Rudman & Phelan, 2008) are inferences of agentic and communal traits. Whereas agency is associated with competence, dominance, and leadership, communion is associated with warmth and caretaking (Abele & Wojciszke, 2007; Eagly et al., 2020; Wiggins, 1991). Inferences of agentic and communal traits are linked to, respectively, prescriptive stereotypes for men and for women (Prentice & Carranza, 2002), and have implications for leadership. Traits traditionally associated with masculinity (e.g., assertiveness) are often *expected* in leaders, whereas traits associated with femininity (e.g., caring) are more closely linked with caregiving roles (Koenig et al., 2011; Rosette & Tost, 2010). Such expectations of agentic and communal inferences can elicit incongruity between the qualities expected of, respectively, effective leaders and of women. Indeed, role congruity theory (Eagly & Karau, 2002) asserts that women displaying agentic behaviors characteristic of leadership are often perceived as deviating from communal expectations, which elicits negative outcomes for them.

One such negative outcome is the backlash that aspiring women leaders often incur for being evaluated as deficient in communal traits (e.g., warmth). People, for instance, justify hiring discrimination against agentic women by making evaluations based on a perceived deficit in social skills (Phelan et al., 2008). Indeed, perceived power-seeking intentions negatively affect people's voting preferences toward hypothetical (Okimoto & Brescoll, 2010) and actual (Gervais & Hillard, 2011) women candidates perceived to be insufficiently communal. Men, by contrast, seem to incur backlash for enjoying feminine hobbies and jobs (Moss-Racusin et al., 2010) that may more broadly reflect low status rather than a perceived lack of agentic traits (Moss-Racusin, 2014).

Inferences of agentic and communal traits are also central to impressions from sexually dimorphic facial features. Here, we define sexually dimorphic facial features as those reflecting prototypically masculinized or feminized sexual

characteristics (see Hu et al., 2018). Although some work has defined facial masculinity and femininity as separable featural constructs (Hester et al., 2021), we treat sexual dimorphism in faces in a continuous way that reflects its treatment in related research (Marcinkowska et al., 2014; Perrett et al., 1998; Rhodes et al., 2003; Wen et al., 2020). Work treating sexually dimorphic facial features in a continuous way has shown across a variety of explicit and implicit tasks that people evaluate masculinized and feminized faces as, respectively, reflecting agentic and communal traits (e.g., Oh et al., 2019; Sutherland et al., 2015; Wen et al., 2020).

Stronger impressions of agentic traits from facial features relate to evaluated leadership suitability (Little, 2014; Re et al., 2013; Spisak et al., 2012). Inferences of power that reflect higher agency among male CEO faces, for example, positively relate to the profits their companies actually enjoy (Rule & Ambady, 2008). Yet the traits that make people successful leaders can vary across domains. Reflecting this assertion, power inferences positively relate to leadership among law executives, whereas social skill inferences reflect leadership in organized crime (Re & Rule, 2017). Trait inferences of agentic and communal qualities reflected in faces thus relate to leadership success depending on how those qualities are valued in specific domains. The trait qualities inferred from faces also differentially relate to leadership success depending on the social category of the target. Indeed, whereas warmth positively relates to judged leadership abilities of Black targets, dominance does so for White targets (Wilson et al., 2017). These findings speak to the importance of target-group membership and target identity in evaluating leadership from faces. We extended this work by examining whether sexually dimorphic facial features that elicit inferences of agentic and communal traits (e.g., Wen et al., 2020) yield differential outcomes for men and women aspiring to leadership.

Supporting this possibility, people often negatively evaluate masculinized relative to feminized women's faces (Carpinella & Johnson, 2016; Gundersen & Kunst, 2019; Hehman et al., 2014; Lick & Johnson, 2014; Oh et al., 2020; Sutherland et al., 2015; but see Rule & Ambady, 2009). People have lower voting intentions, for example, toward women candidates with gender-atypical facial features (Hehman et al., 2014). This replicable pattern (e.g., Cassidy & Liebenow, 2021; Conroy et al., 2020; Ditonto & Mattes, 2018; M. C. Schneider et al., 2021) suggests that women candidates with masculinized (versus feminized) features face an additional barrier to achieving their goals (see Carpinella & Johnson, 2016). Indeed, feminine perceptions of Kamala Harris's face related to more favorable evaluations of her candidacy when controlling for general gender stereotypicality evaluations (Cassidy & Liebenow, 2021).

A direct link showing that sexually dimorphic facial features pose a barrier to women aspiring to leadership roles, however, has yet to be established. For example, much work examining outcomes for aspiring women leaders varying in sexually dimorphic facial features has not compared these women against aspiring men leaders (e.g., Cassidy & Liebenow, 2021). Thus, we cannot discern whether these features are a barrier unique to women rather than a more general one. Other work has shown that people more negatively evaluate women's (versus men's) faces when they appear gender atypical versus typical (Sutherland et al., 2015).

However, this work has done so without determining whether these evaluations extend to differential outcomes for men and women. Without these outcomes, we cannot discern whether patterns from backlash research (Moss-Racusin, 2014; Rudman & Glick, 1999; Rudman & Phelan, 2008) are reflected in evaluations of aspiring leaders who vary in sexually dimorphic facial features.

The current work addressed these gaps in the literature by examining evaluations of college-aged men and women who varied in sexually dimorphic facial features who were ostensibly aspiring to a student leadership role. Much backlash research focuses on domains where gender disparities are well recognized (e.g., politics; Bosak & Sczesny, 2011; Eagly & Karau, 2002; Gervais & Hillard, 2011). We focused on a younger educational setting because women who aspire to leadership roles in domains like politics have often done so over several decades (e.g., Gervais & Hillard, 2011). Obtaining leadership positions as a college-aged adult could create a foundation for the belonging that enables women's aspirations (Belanger et al., 2020; Goodale et al., 2018). Identifying whether nonverbal routes reflective of backlash emerge where such foundations are set can more comprehensively characterize gender disparity in leadership.

Indirectly supporting this premise, recent work showed that STEM professors with more trustworthy (versus dominant) faces are perceived as valuing communal goals and as affording more opportunities to their research groups (Joshi et al., 2023). The current research builds on this finding in two ways. First, we broadened the traits reflective of agency and communion on which participants evaluated faces. Trustworthiness and dominance are primary dimensions in face perception (Oosterhof & Todorov, 2008) that to some extent are, respectively, communal and agentic inferences (Oliveira et al., 2020). Warmth and competence, however, are core dimensions of social perception that broadly reflect stereotypical content about different groups (Cuddy et al., 2008). Moreover, these core dimensions constitute a foundation of gendered thinking pervasive over the lifespan (Martin & Slepian, 2020) that may be changing to some extent. Indeed, whereas the agentic inference of dominance is still traditionally expected of men, both men and women are now expected to be similarly competent (Eagly et al., 2020). By focusing the current research on how inferences of warmth, competence, and dominance vary by sexually dimorphic facial characteristics, we forge a stronger theoretical link to the large body of work on gender stereotypes. Using these traits also forges a link to the literature on role congruity and backlash, which often uses these traits as key representations of communality and agency (Abele et al., 2016; Fiske et al., 2002, 2007; Imhoff et al., 2013; Wen et al., 2020).

Second, we focus on targets *before* they obtain leadership positions versus targets already enjoying them. Although affordances from STEM professor appearance may signal a lab's communal or agentic culture (Joshi et al., 2023), these affordances are made of targets who have already achieved a leadership role. Focusing on targets before leadership is attained is important to test whether sexually dimorphic features may function as a nonverbal *barrier* to leadership rather than a set of features that elicit inferences from established leaders. Notably, feminine appearance can (inaccurately) signal that women (but not men) are not well-suited

for STEM faculty positions (Banchefsky et al., 2016). Such work suggests that sexually dimorphic facial features have the potential to invalidate women's relative to men's leadership aspirations to a greater extent.

Validating work showing that masculinized versus feminized facial features elicit stronger inferences of agentic traits (Walker & Wänke, 2017; Wen et al., 2020), we hypothesized that masculinized relative to feminized faces would be perceived as more competent and dominant, but less warm. Consistent with gender stereotypes potentially most attributed to White individuals (Bruckmüller & Abele, 2013; Johnson et al., 2012), we expected that men's relative to women's faces would be perceived as more competent and dominant, but less warm. Critically, we expected differential evaluative outcomes for men and women based on sexually dimorphic facial features. Consistent with prior backlash research, we expected that women with masculinized versus feminized facial features would receive more negative leadership-related evaluations. Because enacting low-status behaviors (Moss-Racusin et al., 2010) and not femininity per se (Little et al., 2011) elicits negative outcomes for men, we did not expect a difference in evaluations for men with masculinized versus feminized facial features.

EXPERIMENT 1

Experiment 1 tested whether sexually dimorphic facial features differentially affect evaluative outcomes for aspiring student leaders by gender. We expected face impressions reflective of agency (competence and dominance) and communality (warmth) to align with gender stereotypes (e.g., Bruckmüller & Abele, 2013) and sexually dimorphic facial features (Walker & Wänke, 2017; Wen et al., 2020). On the basis of work showing that gender atypicality is negatively related to evaluations of women and not men (Sutherland et al., 2015), we expected more negative outcomes based on sexually dimorphic facial features only for women aspiring to leadership.

METHOD

Participants

In light of recent work (e.g., Wen et al., 2020), we recruited 200 MTurk participants. Twenty-seven failed at least one attention check (described below), yielding 173 analyzed participants ($M_{\text{age}} = 39.61$ years, $SD = 11.88$; $M_{\text{years of education}} = 15.33$, $SD = 2.34$; 69 identifying as female). Of these participants, 135 identified as White, 17 as Black, 14 as Asian, 4 as multiracial, 2 as American Indian/Alaska Native, and 1 as Native Hawaiian/other Pacific Islander. Of these participants, 157 identified as non-Hispanic. All experiments were approved by the UNC Greensboro IRB. All participants provided informed consent. To benefit future work, we provide sensitivity analyses for interactive effects between sexually dimorphic facial features and target gender on outcomes across experiments (see below).

Data and analytic code are available on the Open Science Framework (https://osf.io/ygwj5/?view_only=2ad8762941c84303a2a5a969b5629aef).

Stimuli

Twenty identities each of neutrally expressive White men and White women used in work examining contributions of sexually dimorphic facial features to social outcomes (e.g., Goh et al., 2022) were selected. Each face identity had masculinized and feminized versions (Figure 1; for details on face manipulation, see DeBruine & Jones, 2017) that have been used in related work (e.g., DeBruine & Jones, 2017; Little et al., 2011; Shiramizu et al., 2024). Moreover, using the same identities across task versions manipulates sexually dimorphic features while holding all other face aspects constant. Masculinized and feminized faces broadly differed in similar ways. Masculinized faces had thicker, lower eyebrows and wider, square jaws. Feminized faces had thinner, higher eyebrows and smaller, round jaws. Relevant here, the faces were all emerging adults and thus appropriate for evaluation as potential college-aged student leaders.

Task

Participants evaluated candidates, saw candidate pictures, and provided impressions. The description read:

The Student Policies Manager is a prestigious position for a student leader who displays strong competency in their work, independence, and motivation to lead their college community. Responsibilities for this position are to advise the board of trustees on student and university issues, lead student council in weekly meetings, and be a competent facilitator of student and faculty queries. To succeed in this role, the candidate should be highly competent, independent, intelligent, and determined. Past leadership experience is desirable. Past students in this position have gone on to work for prestigious companies and even been elected for local, city, and state legislative seats.

A one-tailed t test against the scale (1 [*very feminine*] to 7 [*very masculine*]) midpoint showed that 40 participants who did not complete the experiments evaluated the description as gender neutral ($M = 3.90$, $SD = 0.22$), $t(38) = -0.51$, $p = .61$, $d = 0.082$.

As attention checks, participants indicated whether the description described a party planner, student leader, restaurant manager, or corporate intern. Nineteen were incorrect and excluded. Participants indicated whether they would evaluate candidates for Student Policies Manager, Event Coordinator, Treasurer, or Social Media Liaison. Eight were incorrect and excluded.

In the main task, participants saw pictures of and evaluated 20 White men and 20 White women in randomly presented trials. Within each gender, 10 faces each

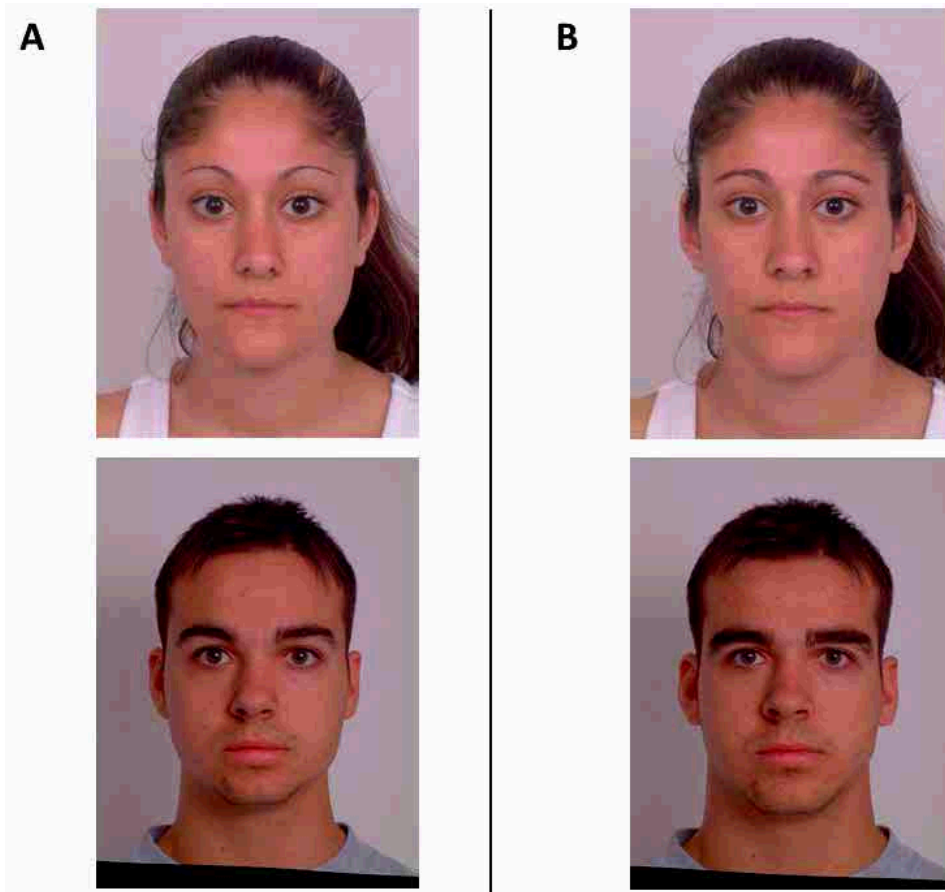


FIGURE 1. Examples of feminized (a) and masculinized (b) face identities.

were masculinized or feminized. Each identity was shown once and was displayed through all evaluations. Two task versions counterbalanced target facial features within-participants. In each self-paced trial, participants rated a candidate on *warmth*, *competence*, and *dominance* using scales ranging from 1 (*not at all*) to 7 (*extremely*). These traits are commonly used in work assessing agency (competence and dominance) and communion (warmth) (Abele et al., 2016; Fiske et al., 2007; Wen et al., 2020). Critically, they have been extensively linked to gender stereotypes (Fiske et al., 2002; Imhoff et al., 2013) and used in related work on impressions from sexually dimorphic facial features (Walker & Wänke, 2017). The General Discussion includes a discussion of potential limitations of using these traits.

On the next screen, participants evaluated the candidate ("How likely are you to endorse this candidate for the role of Student Policies Manager?"; "How likely is this candidate to do well as Student Policies Manager?"; "How likely are other

people to like reporting to this candidate as the Student Policies Manager?”) using scales ranging from 1 (*not at all likely*) to 7 (*extremely likely*).

Participants then indicated how well they adhered to task instructions and paid attention on a scale from 1 (*not at all*) to 7 (*completely*), $M = 6.82$, $SD = 0.66$.

Characterization Measures

We characterized participants on measures collected in a random order (see Table 1 for descriptive statistics and intercorrelations). Participants indicated political ideology over four items (overall, economic issues, social issues, and foreign policy issues) on a scale ranging from 1 (*extremely conservative*) to 9 (*extremely liberal*). Responses (Cronbach's $\alpha = .95$) were averaged. Participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). The ASI measures hostile (e.g., “Women are too easily offended”) and benevolent (e.g., “Women should be cherished and protected by men”) sexism on a scale ranging from 0 (*disagree strongly*) to 5 (*agree strongly*). Higher scores indicate more sexism. Items measuring hostile (Cronbach's $\alpha = .91$) and benevolent (Cronbach's $\alpha = .89$) sexism were averaged. Participants completed the Social Roles Questionnaire (SRQ; Baber & Tucker, 2006) to measure traditional social role beliefs. Participants indicated how much they agreed with each statement (e.g., “For many important jobs, it is better to choose men instead of women”) from 0% to 100% in 10% increments. Higher scores reflect more traditional social role beliefs. Items (Cronbach's $\alpha = .90$) were averaged. Participants completed a social dominance orientation questionnaire (Ho et al., 2015) using a scale ranging from 1 (*strongly oppose*) to 7 (*strongly favor*). We averaged responses (e.g., “An ideal society requires some groups to be on top and others to be on the bottom.”; Cronbach's $\alpha = .91$).

Data Analyses

We fitted mixed-effects models using *lme4* (Bates et al., 2015) in R, estimating confidence intervals using the *confint* function. *P* values were calculated using *lmerTest* (Kuznetsova et al., 2017). Estimated marginal means were obtained using *emmeans* (Lenth et al., 2018). The use of “*z*” in the reported inferential statistics reflects contrasts from *emmeans* to characterize interactions from multilevel models. We used the *emtrends* function within *emmeans* to characterize simple slopes in interactions including continuous variables.

RESULTS

Gender Differences on Characterization Measures

Men and women participants had similar hostile sexism, benevolent sexism, traditional social role beliefs, social dominance orientation, political ideology, age, and years of education (Table 2a).

TABLE 1. Intercorrelations for Participant Characterization Measures (Means, Standard Deviations, and Correlations With Confidence Intervals) for Experiments 1 (Below Diagonal) and 2 (Above Diagonal)

Variable	Experiment 1 Experiment 2		1	2	3	4	5
	<i>M (SD)</i>	<i>M (SD)</i>					
1. Ideology	4.42 (1.73)	4.59 (1.74)		-.43** [-.51, -.34]	-.32** [-.41, -.23]	-.43** [-.51, -.34]	-.46** [-.53, -.37]
2. Hostile Sexism	1.78 (1.19)	1.87 (1.25)	-.35** [-.47, -.21]		.50** [.42, .58]	.74** [.69, .79]	.66** [.60, .71]
3. Benevolent Sexism	2.07 (1.12)	2.09 (1.10)	-.28** [-.41, -.14]	.51** [.39, .61]		.67** [.61, .72]	.38** [.29, .46]
4. Social Roles Questionnaire	33.81 (19.93)	35.74 (21.03)	-.36** [-.48, -.22]	.73** [.66, .80]	.67** [.58, .74]		.64** [.58, .70]
5. Social Dominance Orientation	2.53 (1.42)	2.61 (1.49)	-.46** [-.57, -.33]	.70** [.61, .77]	.42** [.29, .53]	.71** [.62, .77]	

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * $p < .05$. ** $p < .01$.

Effects of Gender and Sexually Dimorphic Facial Features on Trait Evaluations

We regressed evaluations on Trait (effects coded with competence as the reference level), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman), and their interactions. We included Perceiver Gender on an exploratory basis. See Table 3 for all coefficient information for the fixed effects. The random effects structure allowed intercepts to vary by participant and face identity and facial characteristics effects to vary by participant and face identity. See Supplemental Materials for main effects and breakdown of each individual trait variable.

Two-Way Interactions

Trait and Face Gender interacted (Figure 2a). Participants evaluated women's (Estimate = 4.46, 95% CI [4.29, 4.63]) relative to men's (Estimate = 3.95, 95% CI [3.78, 4.12]) faces as more warm, $b = 0.51$, $SE = 0.09$, $z = 5.62$, $p < .01$. Participants evaluated women's (Estimate = 4.75, 95% CI [4.58, 4.93]) relative to men's (Estimate = 4.30, 95% CI [4.13, 4.47]) faces as more competent, $b = 0.45$, $SE = 0.09$, $z = 4.99$, $p < .001$. In contrast, participants evaluated women's (Estimate = 4.35,

TABLE 2. Descriptive (*M* [*SD*]) and Inferential Statistics for Participant Characterization Measures in Experiments 1 (a) and 2 (b)

Experiment 1	Men (<i>N</i> = 104)	Women (<i>N</i> = 69)	<i>t</i>	<i>p</i>	<i>d</i> [95% CI]
Hostile sexism	1.87 (1.17)	1.63 (1.22)	1.33	.18	0.21 [−0.12, 0.61]
Benevolent sexism	2.05 (1.10)	2.09 (1.17)	−0.28	.78	0.04 [−0.39, 0.30]
Traditional gender role beliefs	33.74 (19.72)	33.92 (20.39)	−0.06	.95	0.01 [−6.31, 5.94]
Social dominance Orientation	2.46 (1.42)	2.63 (1.42)	−0.75	.45	0.12 [−0.60, 0.27]
Political ideology	4.38 (1.77)	4.49 (1.68)	−0.42	.68	0.07 [−0.64, 0.42]
Age	39.84 (10.43)	39.26 (11.19)	0.35	.73	0.05 [−2.72, 3.87]
Years of education	15.34 (2.06)	15.32 (2.94)	0.05	.96	0.01 [−0.73, 0.77]
Experiment 2	Men (<i>N</i> = 209)	Women (<i>N</i> = 164)	<i>t</i>	<i>p</i>	<i>d</i> [95% CI]
Hostile sexism	2.15 (1.18)	1.51 (1.24)	5.10	< .001	0.53 [0.39, 0.89]
Benevolent sexism	2.20 (1.04)	1.96 (1.17)	2.10	.04	0.22 [0.02, 0.47]
Traditional gender role beliefs	38.80 (19.88)	31.85 (21.86)	3.20	.001	0.33 [2.68, 11.21]
Social dominance Orientation	2.77 (1.44)	2.41 (1.53)	2.36	.02	0.25 [0.06, 0.67]
Political ideology	4.49 (1.66)	4.73 (1.83)	−1.32	.19	0.14 [−0.60, 0.12]
Age	39.96 (11.30)	41.76(12.99)	−1.43	.15	0.15 [−4.28, 0.68]
Years of education	15.42 (2.77)	15.38 (2.21)	0.17	.87	0.02 [−0.48, 0.57]

95% CI [4.18, 4.52]) and men's (Estimate = 4.44, 95% CI [4.27, 4.61]) faces similarly on dominance, $b = -0.09$, $SE = 0.09$, $z = -0.96$, $p = .93$.

Trait and Facial Characteristics interacted (Figure 2b). Participants evaluated feminized (Estimate = 4.34, 95% CI [4.19, 4.49]) relative to masculinized (Estimate = 4.08, 95% CI [3.93, 4.23]) faces as more warm, $b = 0.26$, $SE = 0.04$, $z = 7.35$, $p < .001$. Participants evaluated feminized (Estimate = 4.60, 95% CI [4.45, 4.75]) relative to masculinized (Estimate = 4.46, 95% CI [4.31, 4.61]) faces as more competent, $b = 0.14$, $SE = 0.04$, $z = 3.89$, $p = .001$. In contrast, participants evaluated feminized (Estimate = 4.28, 95% CI [4.13, 4.43]) relative to masculinized (Estimate = 4.51, 95% CI [4.36, 4.66]) faces as less dominant, $b = -0.23$, $SE = 0.04$, $z = -6.65$, $p < .001$.

Facial Characteristics and Face Gender interacted (Figure 3a). Participants evaluated feminized (Estimate = 4.61, 95% CI [4.44, 4.78]) versus masculinized (Estimate = 4.44, 95% CI [4.27, 4.61]) women's faces higher in traits overall, $b = 0.17$, $SE = 0.03$, $z = 4.99$, $p < .001$. In contrast, participants similarly evaluated traits in feminized (Estimate = 4.20, 95% CI [4.03, 4.37]) versus masculinized (Estimate = 4.26, 95% CI [4.09, 4.43]) men's faces, $b = -0.06$, $SE = 0.03$, $z = -1.82$, $p = .26$.

Face Gender and Perceiver Gender interacted. Men evaluated women's (Estimate = 4.43, 95% CI [4.24, 4.62]) versus men's (Estimate = 4.08, 95% CI [3.89, 4.28]) faces as higher in traits, $b = 0.34$, $SE = 0.09$, $z = 3.93$, $p < .001$. To a lesser extent,

TABLE 3. Effects of Face Gender, Facial Characteristics, and Participant Gender on Evaluated Warmth (a), Competence (b), and Dominance (c) in Experiment 1

Predictor	a. Warmth			b. Competence			c. Dominance		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(Intercept)	4.21	[4.03, 4.39]	< .001	4.53	[4.36, 4.69]	< .001	4.39	[4.23, 4.56]	< .001
Face Gender	0.26	[0.13, 0.38]	< .001	0.23	[0.12, 0.34]	< .001	-0.04	[-0.14, 0.05]	.381
Facial Characteristics	0.12	[0.08, 0.17]	< .001	0.07	[0.03, 0.10]	< .001	-0.11	[-0.15, -0.07]	< .001
Participant Gender	0.10	[-0.03, 0.23]	.127	0.09	[-0.03, 0.21]	.156	0.16	[0.03, 0.29]	.017
Face Gender × Facial Characteristics	0.03	[-0.01, 0.07]	.147	0.04	[0.00, 0.07]	.025	0.11	[0.08, 0.14]	< .001
Face Gender × Participant Gender	-0.07	[-0.09, -0.04]	< .001	-0.04	[-0.07, -0.02]	.002	0.02	[-0.01, 0.05]	.143
Facial Characteristics × Participant Gender	-0.02	[-0.05, 0.01]	.178	-0.01	[-0.04, 0.02]	.588	0.02	[-0.02, 0.05]	.386

Note. Participant Gender: (-1 = men participants, 1 = women participants); Face Gender: (-1 = men's faces, 1 = women's faces); Facial Characteristics: (-1 = masculine, 1 = feminine).

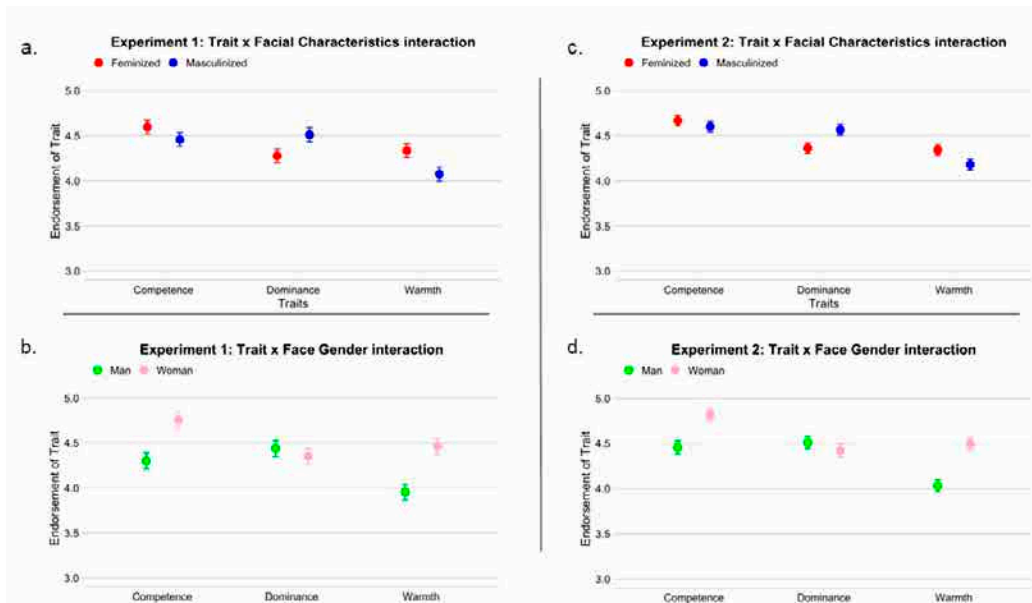


FIGURE 2. Across Experiments 1(a) and 2(c), interactions between trait and facial characteristics emerged on trait endorsements. Participants had stronger trait endorsements of competence and warmth for feminized versus masculinized faces, and stronger trait endorsements of dominance for masculinized versus feminized faces. Across Experiments 1(b) and 2(d), interactions between trait and face gender emerged on trait endorsements. Participants had stronger trait endorsements of competence and warmth for women's versus men's faces, and stronger trait endorsements of dominance for men's versus women's faces.

women evaluated women's (Estimate = 4.61, 95% CI [4.39, 4.83]) versus men's (Estimate = 4.38, 95% CI [4.16, 4.60]) faces as higher in Traits, $b = 0.24$, $SE = 0.09$, $z = 2.62$, $p = .04$.

Three-Way Interactions

Trait, Facial Characteristics, and Face Gender. Trait qualified the interaction between Facial Characteristics and Face Gender (Figure 4a). We characterized this interaction by examining the Facial Characteristics effect for each Face Gender at each Trait.

For women candidates, participants evaluated feminized (Estimate = 4.62, 95% CI [4.44, 4.80]) versus masculinized (Estimate = 4.30, 95% CI [4.13, 4.48]) faces as more warm, $b = 0.32$, $SE = 0.05$, $z = 6.67$, $p < .001$. For men candidates, participants evaluated feminized (Estimate = 4.05, 95% CI [3.87, 4.23]) versus masculinized (Estimate = 3.85, 95% CI [3.68, 4.03]) faces as more warm, $b = 0.20$, $SE = 0.05$, $z = 4.14$, $p = .002$.

For women candidates, participants evaluated feminized (Estimate = 4.86, 95% CI [4.68, 5.04]) versus masculinized (Estimate = 4.65, 95% CI [4.47, 4.83]) faces as

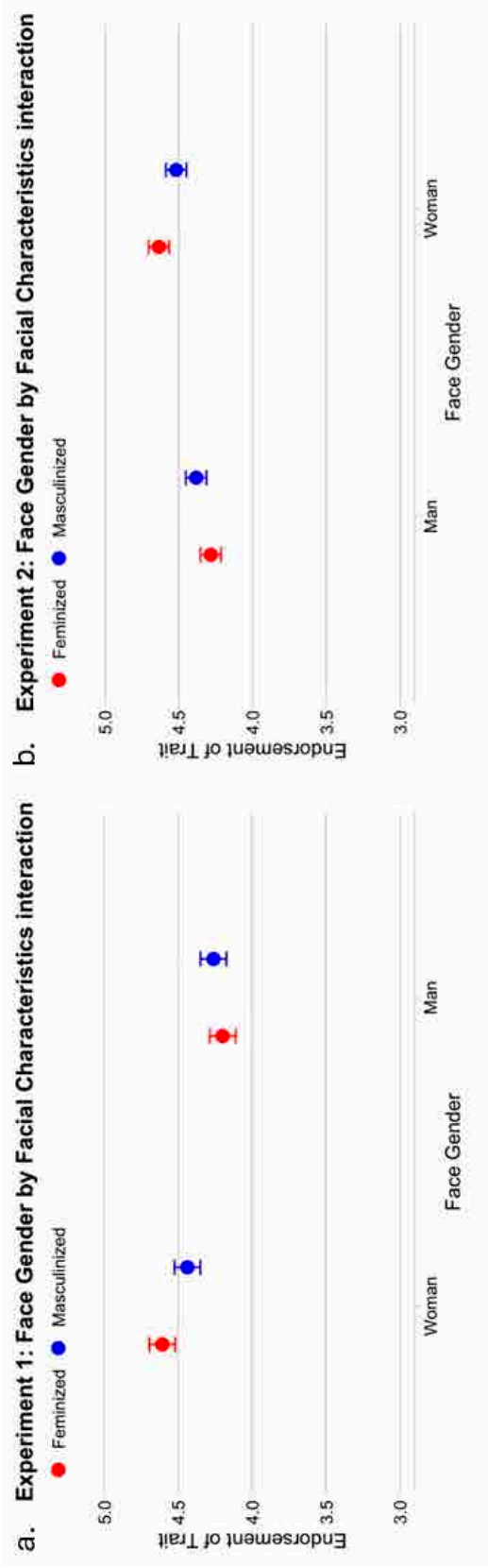


FIGURE 3. Across Experiments 1(a) and 2(b), interactions between face gender and facial characteristics emerged on trait endorsements. Participants had stronger trait endorsements of women's faces that were feminized versus masculinized. For men's faces, trait endorsements did not differ by facial characteristics.

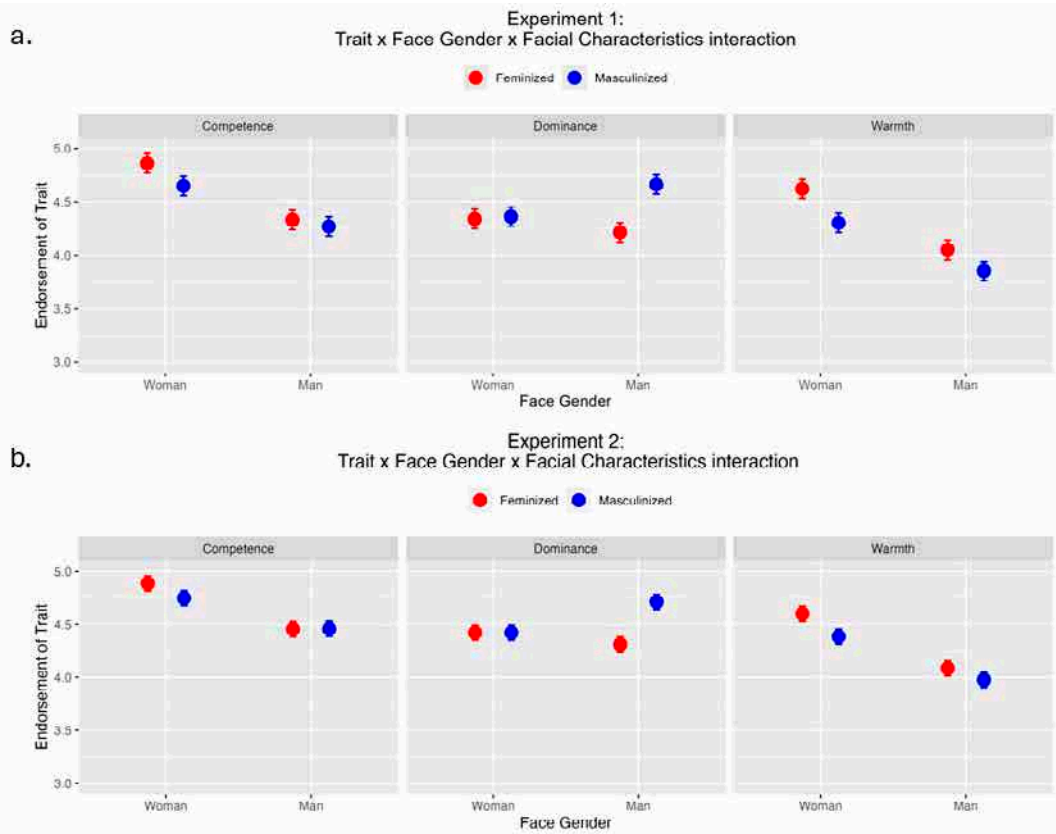


FIGURE 4. Across both Experiment 1(a) and Experiment 2(b), Trait qualified the interaction between Facial Characteristics and Face Gender. For women candidates, participants were more likely to evaluate feminized women candidates, relative to masculinized women candidates, as more warm and more competent. For men candidates, there was no significant difference between participants' evaluations of both warmth and competency for feminized men candidates relative to masculinized men candidates. However, for dominance, participants were more likely to evaluate masculinized men candidates, relative to feminized men candidates, as more dominant. No difference emerged between feminized and masculinized women candidates for dominance evaluations.

more competent, $b = 0.21$, $SE = 0.05$, $z = 4.42$, $p < .001$. For men candidates, in contrast, participants evaluated feminized (Estimate = 4.33, 95% CI [4.15, 4.51]) versus masculinized (Estimate = 4.27, 95% CI [4.09, 4.45]) faces as similarly competent, $b = 0.06$, $SE = 0.05$, $z = 1.14$, $p = .98$.

For women candidates, people evaluated feminized (Estimate = 4.34, 95% CI [4.16, 4.52]) and masculinized (Estimate = 4.36, 95% CI [4.18, 4.54]) faces as similarly dominant, $b = -0.02$, $SE = 0.05$, $z = -0.46$, $p > .99$. For men candidates, in contrast, participants evaluated masculinized (Estimate = 4.66, 95% CI [4.48, 4.84]) versus feminized (Estimate = 4.22, 95% CI [4.04, 4.39]) faces as more dominant, $b = -0.45$, $SE = 0.05$, $z = -9.32$, $p < .001$.

Trait, Face Gender, and Perceiver Gender. Trait qualified the interaction between Face Gender and Perceiver Gender (Figure 5a). We characterized this interaction by examining how evaluations of Traits differed for each Face and Perceiver Gender.

Men evaluated men's faces as more dominant (Estimate = 4.30, 95% CI [4.10, 4.49]) than competent (Estimate = 4.17, 95% CI [3.97, 4.37]), $b = -0.13$, $SE = 0.04$, $z = -3.515$, $p = .02$. In contrast, they evaluated women's faces as more competent (Estimate = 4.71, 95% CI [4.51, 4.90]) than dominant (Estimate = 4.17, 95% CI [3.97, 4.37]), $b = 0.54$, $SE = 0.04$, $z = 14.49$, $p < .001$. Men evaluated men's faces as more competent (Estimate = 4.17, 95% CI [3.97, 4.37]) than warm (Estimate = 3.79, 95% CI [3.59, 3.98]), $b = 0.38$, $SE = 0.04$, $z = 10.36$, $p < .001$. They also evaluated women's faces as more competent (Estimate = 4.71, 95% CI [4.51, 4.90]) than warm (Estimate = 4.42, 95% CI [4.23, 4.62]), $b = 0.28$, $SE = 0.04$, $z = 7.63$, $p < .001$. Men evaluated men's faces as more dominant (Estimate = 4.30, 95% CI [4.10, 4.49]) than warm (Estimate = 3.79, 95% CI [3.59, 3.98]), $b = 0.51$, $SE = 0.04$, $z = 13.88$, $p < .001$. In contrast, they rated women's faces as more warm (Estimate = 4.42, 95% CI [4.23, 4.62]) than dominant (Estimate = 4.17, 95% CI [3.97, 4.37]), $b = -0.25$, $SE = 0.04$, $z = -6.86$, $p < .001$.

Unlike men, women similarly evaluated men's faces in dominance (Estimate = 4.58, 95% CI [4.35, 4.80]) and competence (Estimate = 4.43, 95% CI [4.21, 4.66]), $b = -0.15$, $SE = 0.05$, $z = -3.23$, $p = .06$. Like men, women evaluated women's faces as more competent (Estimate = 4.80, 95% CI [3.97, 4.37]) than dominant (Estimate = 4.53, 95% CI [4.31, 4.76]), $b = 0.27$, $SE = 0.05$, $z = 5.99$, $p < .001$. Like men, women evaluated men's faces as more competent (Estimate = 4.43, 95% CI [4.21, 4.66]) than warm (Estimate = 4.12, 95% CI [3.89, 4.34]), $b = 0.31$, $SE = 0.05$, $z = 6.91$, $p < .001$. Also, like men, women evaluated women's faces as more competent (Estimate = 4.80, 95% CI [3.97, 4.37]) than warm (Estimate = 4.50, 95% CI [4.28, 4.73]), $b = 0.30$, $SE = 0.05$, $z = 6.65$, $p < .001$. Like men, women evaluated men's faces as more dominant (Estimate = 4.58, 95% CI [4.35, 4.80]) than warm (Estimate = 4.12, 95% CI [3.89, 4.34]), $b = 0.46$, $SE = 0.05$, $z = 10.13$, $p < .001$. Unlike men, women rated women's faces as similarly warm (Estimate = 4.50, 95% CI [4.28, 4.73]) and dominant (Estimate = 4.53, 95% CI [4.31, 4.76]), $b = 0.03$, $SE = 0.05$, $z = 0.66$, $p > .99$.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

Because responses to the three outcome evaluations were reliable (Cronbach's $\alpha = .93$), we created a composite outcome rating by averaging across outcomes. We regressed these composite ratings on Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman), and their interactions. Perceiver Gender was included on an exploratory basis. The model included a random effects structure that allowed intercepts to vary by participant and face identity and that allowed Facial Characteristics effects to vary by participant and face identity. See Table 4 for all model coefficient

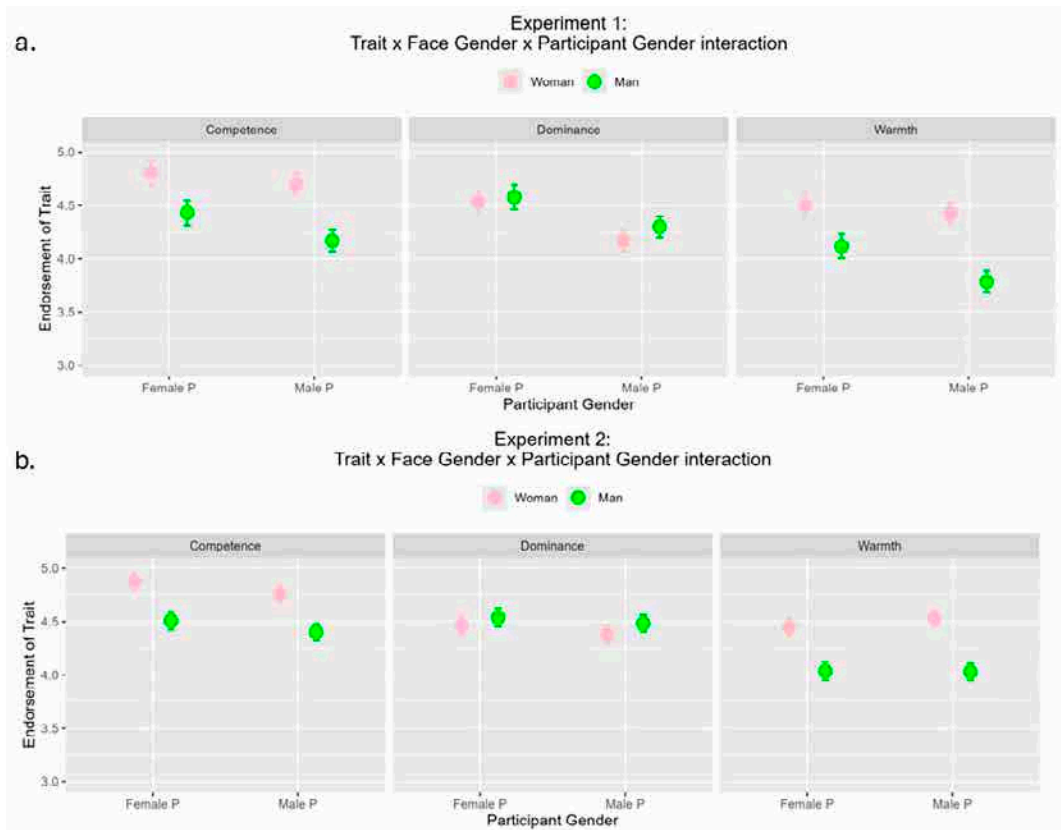


FIGURE 5. In Experiment 1(a) and Experiment 2(b), Trait qualified the interaction between Participant Gender and Face Gender. Both men and women participants evaluated men's faces higher in dominance than in competence. In contrast, both men and women participants evaluated women's faces as higher in competence than in dominance. Both men and women participants evaluated men's and women's faces as higher in competence than warmth. Both men and women participants evaluated men's faces as higher in dominance than in warmth, while women participants rated women's faces as equally dominant and warm.

information. See Supplemental Materials for main effects and breakdown of each individual outcome variable.

Two-Way Interactions

As hypothesized, Face Gender and Facial Characteristics interacted (Figure 6a). Participants endorsed feminized (Estimate = 4.55, 95% CI [4.33, 4.76]) more than masculinized (Estimate = 4.31, 95% CI [4.10, 4.52]) women's faces, $b = 0.24$, $SE = 0.05$, $z = 4.74$, $p < .001$. No difference between feminized (Estimate = 3.98, 95% CI [3.76, 4.19]) versus masculinized (Estimate = 3.94, 95% CI [3.73, 4.15]) men's faces emerged, $b = 0.03$, $SE = 0.05$, $z = 0.68$, $p = .90$. Using our specified model

TABLE 4. Effects of Face Gender, Facial Characteristics, and Participant Gender on (a) Candidate Endorsement, (b) Likelihood of Doing Well in Position, and (c) People Liking Candidate as a Leader in Experiment 1

Predictor	Outcome 1			Outcome 2			Outcome 3		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(Intercept)	4.11	[3.92, 4.29]	< .001	4.26	[4.09, 4.43]	< .001	4.22	[4.04, 4.39]	< .001
Face Gender	0.26	[0.14, 0.38]	< .001	0.25	[0.13, 0.36]	< .001	0.20	[0.09, 0.31]	< .001
Facial Characteristics	0.07	[0.03, 0.11]	< .001	0.07	[0.03, 0.10]	< .001	0.07	[0.03, 0.11]	.001
Participant Gender	0.12	[−0.03, 0.26]	.108	0.12	[−0.00, 0.25]	.054	0.14	[0.00, 0.27]	.045
Face Gender × Facial Characteristics	0.05	[0.01, 0.08]	.016	0.03	[0.00, 0.07]	.045	0.07	[0.03, 0.11]	< .001
Face Gender × Participant Gender	−0.04	[−0.07, −0.01]	.005	−0.04	[−0.07, −0.01]	.006	−0.06	[−0.09, −0.03]	< .001
Facial Characteristics × Participant Gender	−0.01	[−0.04, 0.01]	.329	−0.00	[−0.03, 0.03]	.995	−0.02	[−0.05, 0.01]	.318

Note. Participant Gender: (−1 = men participants; 1 = women participants); Face Gender: (−1 = men's faces; 1 = women's faces); Facial Characteristics: (−1 = masculine, 1 = feminine).

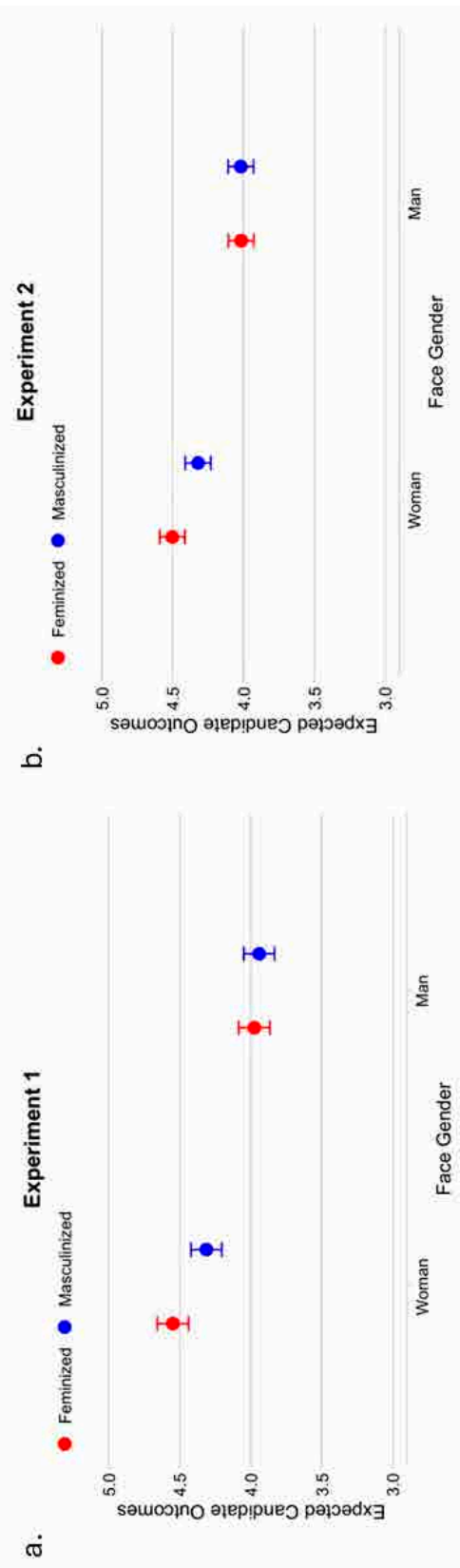


FIGURE 6. Across Experiments 1(a) and 2(b), interactions between face gender and facial characteristics emerged on evaluative outcomes. Participants had more positive evaluative outcomes for women candidates when they had feminized versus masculinized facial characteristics. For men candidates, evaluative outcomes did not differ by facial characteristics.

and sample size, a sensitivity analysis revealed a minimum detectable interaction effect size of 0.0605 with 80% power and $\alpha = .05$.

Face Gender and Perceiver Gender interacted. Men endorsed women (Estimate = 4.35, 95% CI [4.12, 4.58]) more than men (Estimate = 3.79, 95% CI [3.56, 4.02]) candidates, $b = 0.56$, $SE = 0.12$, $z = 4.67$, $p < .001$. To a lesser extent, women endorsed women (Estimate = 4.51, 95% CI [4.25, 4.77]) more than men (Estimate = 4.13, 95% CI [3.87, 4.39]) candidates, $b = 0.38$, $SE = 0.12$, $z = 3.12$, $p = .01$.

DISCUSSION

Warmth and dominance impressions varied by sexually dimorphic facial features and aligned with prior work (Imhoff et al., 2013; Oh et al., 2019; Wen et al., 2020). Masculinized faces were more dominant than feminized faces, whereas feminized faces were warmer than masculinized faces. Impressions of men's and women's faces were partially consistent with gender stereotypes (Prentice & Carranza, 2002). Consistent with stereotypes, women's versus men's faces were warmer.

Competence and dominance impressions were inconsistent with stereotypes. First, women's versus men's faces were perceived to be more competent. Furthermore, women's competence was strengthened by gender stereotypic features, but men's were not. Because women versus men comprise more of college populations (Bauman & Cranney, 2020), one possibility is that people perceive college-aged women as especially scholarly, reflecting increasing perceptions of women's competence over time (Eagly et al., 2020). That feminized versus masculinized faces were perceived as more competent supports this possibility. Second, women's and men's faces were perceived as similarly dominant. Furthermore, men's evaluated dominance was strengthened by having gender stereotypic facial features, but women's evaluated dominance was not. Indeed, there are category and sexually dimorphic features that could be additive effects in the extent to which they reflect gender stereotypic traits of the gender category. This is consistent with work showing that prototypical members of the category are more stereotypically evaluated (e.g., faces evaluated as stereotypically Black are perceived negatively relative to less stereotypical faces; Kleider-Offutt et al., 2021).

Evaluations both complemented and contrasted backlash research. In complement (e.g., Rudman, 1998), only women with masculinized relative to feminized faces were less favorably evaluated. This pattern supports sexually dimorphic facial features as a route to backlash against women. It also suggests an effect emergent in settings involving college-aged students, an understudied population in backlash research.

Women (versus men) candidates and feminized (versus masculinized) faces each received more favorable outcomes. In fact, feminized women's faces received the most favorable outcomes when compared to all other combinations of gender and facial characteristics (see Supplemental Materials). These patterns may seem surprising given work showing that women versus men are less favorably evaluated when aspiring to leadership (Eagly & Karau, 2002). Leadership is traditionally

aligned with traits better reflecting masculinity (e.g., power) than femininity (e.g., warmth; Koenig et al., 2011; Rosette & Tost, 2010). Indeed, people are quicker to associate agentic behaviors with leadership when the targets in question are men versus women (Scott & Brown, 2006). Here, women were evaluated as agentic via their higher competence, but were not penalized overall relative to men.

We offer speculation on why these women and feminized faces had more favorable outcomes in the current work. First, women being preferred over men could be because the leadership position may not have been perceived as high status. Because the leadership role was for students, it could be that the role was not perceived to be as serious as a leadership role in other stages of life or in other contexts (e.g., academic faculty, politics; Joshi et al., 2023; B. Schneider et al., 2016; M. C. Schneider, 2021). People might be more willing to endorse women for leadership roles when such roles are perceived to be temporary positions or lower in status and importance, such as student positions and leadership positions while still enrolled at school versus in the workplace. Indeed, gender disparities in support for leadership are smaller for leadership positions where a leader is perceived to have less versus more autonomy in decision-making (Smith et al., 2007).

A second possibility emerges from using an educational domain. Because people believe schools embody feminine values and practices, they expect women to behave in belief-consistent ways (Yee, 1973). Women's overrepresentation (Bauman & Cranney, 2020) and gendered beliefs about schools and college students may create a context where women are perceived as more appropriate for leadership. Indeed, meta-analysis data have shown that gender stereotypes shaped people's perceptions of leaders as less masculine in educational organizations compared to other domains (Koenig et al., 2011). In general, people believe that women relative to men are better suited for stereotypically feminine jobs (e.g., Gadassi & Gati, 2009). Nevertheless, masculinized features negatively affected evaluations of these women. Even when women appear to be advantaged, the interactive effect depicts a *limited* advantage. Given a disproportionate weight placed on women's physical appearance (Jackson et al., 1987; Quittkat et al., 2019), this limited advantage may have enduring consequences for aspiring women leaders.

A third possibility regards the job description potentially being more interpersonal than analytical. If true, the gender difference may reflect women being seen as more appropriate for interpersonal roles than men (e.g., Gadassi & Gati, 2009). Interestingly, gender differences in evaluations and competence attributions were stronger for men versus women perceivers. This pattern counters a possibility that women being more favorably evaluated stemmed from in-group favoritism (e.g., Rudman & Goodwin, 2004). Speculatively, men could be more likely to endorse women for leadership in contexts that align with gender roles (e.g., Banchevsky et al., 2016).

We designed Experiment 2 to better understand overall gender effects. Changes allowed us to replicate the effects of Experiment 1 and to address whether gendered beliefs about college students, the gendered nature of the job, or both affected a tendency to endorse women over men for the student leadership position.

EXPERIMENT 2

Experiment 2 served as replication of Experiment 1 and explored why women were favored over men for the student leadership position. We changed Experiment 1 in two ways. First, we manipulated the gendered nature of the position. This manipulation allowed us to assess whether women are evaluated more favorably than men only when the position is stereotypically feminine. Second, we assessed whether gender-biased beliefs related to gender differences in outcomes. Indeed, prior work discerns that previously held negative beliefs and stereotypes about a gender group can influence people's judgments and evaluations of that group in real life (Bligh et al., 2012; Liebenow et al., 2022; Oh et al., 2019; Prentice & Carranza, 2002; Smith et al., 2007). Here, we measured participants' beliefs about college-aged men's and women's warmth, competence, dominance, and responsibility. We speculated that college women are perceived as more competent *compared to college men* overall, and that people may prefer college-aged women versus men in a student leadership position. Therefore, in the present work, we explored the possibility that if people believe that college women versus college men are more warm, competent, dominant, or responsible, participants should also more favorably evaluate college women candidates over college men candidates for the Student Policies Manager position.

METHOD

Participants

Given the between-participants manipulation, we doubled the sample from Experiment 1 to 400 MTurk participants. Eighteen failed the first attention check, and nine failed the second, yielding 373 analyzed participants ($M_{\text{age}} = 40.75$ years, $SD = 13.45$; $M_{\text{years of education}} = 15.41$, $SD = 2.54$, 164 identifying as female). Of these participants, 296 identified as White, 35 as Black, 29 as Asian, 8 as multiracial, 3 as Unknown, and 2 as American Indian/Alaska Native. Of these participants, 351 also identified as non-Hispanic.

Task

Experiment 2 replicated Experiment 1 with the following changes. Participants were randomly assigned to read an analytical ($n = 199$) or an interpersonal ($n = 174$) description. Men and women participants were evenly distributed across versions, $\chi^2(1) < .001$, $p = .99$.

The analytical description read:

The Student Policies Manager is a prestigious position for a student leader who displays a strong work ethic, determination, and persistence to lead their college community. Responsibilities for this position are to manage the board of trustees on student and university issues, oversee the student council in weekly meetings, and be highly competent when facilitating queries from students and faculty. To succeed in this role, the candidate should be decisive, self-sufficient, outspoken,

and authoritative. Past leadership experience is desirable. Past students in this position have gone on to work for prestigious companies and even been elected for local, city, and state legislative seats.

The interpersonal description read:

The Student Policies Manager is a prestigious position for a student leader who displays a strong community-oriented approach to their work, inclusivity, and cooperation to lead their college community. Responsibilities for this position are to support the board of trustees on student and university issues, guide the student council in weekly meetings, and create a communal environment to facilitate student and faculty queries. To succeed in this role, the candidate should be nurturing, empathetic, responsive, and trustworthy. Past leadership experience is desirable. Past students in this position have gone on to work for prestigious companies and even been elected for local, city, and state legislative seats.

Forty participants who did not complete the experiments rated descriptions on their interpersonal and analytical nature, valence, and gender stereotypicality using scales from 1 (*not at all*) to 7 (*very*). The analytical job was more than moderately analytical ($M = 5.67$, $SD = 1.52$), $t(38) = 7.98$, $p < .001$, $d = 1.28$. The interpersonal job was more than moderately interpersonal ($M = 5.92$, $SD = 1.44$), $t(38) = 8.34$, $p < .001$, $d = 1.34$. The interpersonal versus analytical job was more interpersonal, $t(38) = 4.62$, $p < .001$, $d = 1.07$. The analytical versus interpersonal job was more analytical, $t(38) = -4.32$, $p < .001$, $d = 0.91$. The interpersonal ($M = 5.31$, $SD = 1.42$) versus analytical ($M = 3.46$, $SD = 1.74$) job was more stereotypically feminine, $t(38) = 4.35$, $p < .001$, Cohen's $d = 1.16$. The interpersonal ($M = 5.54$, $SD = 1.27$) and analytical ($M = 5.13$, $SD = 1.54$) jobs had similar valence, $t(38) = 1.65$, $p = .11$, $d = 0.29$.

Four versions counterbalanced job description (analytical/interpersonal) between-participants and facial features (masculinized/feminized) within-participants. Participants again indicated that they stayed on task and followed instructions ($M = 6.82$, $SD = 0.47$).

Beliefs About College Students

After the main task, participants indicated beliefs about men and women college students' warmth ($M_{\text{women}} = 5.14$, $SD_{\text{women}} = 1.13$; $M_{\text{men}} = 4.37$, $SD_{\text{men}} = 1.43$), competence ($M_{\text{women}} = 5.45$, $SD_{\text{women}} = 1.24$; $M_{\text{men}} = 5.04$, $SD_{\text{men}} = 1.24$), dominance ($M_{\text{women}} = 4.36$, $SD_{\text{women}} = 1.32$; $M_{\text{men}} = 5.28$, $SD_{\text{men}} = 1.20$), and responsibility ($M_{\text{women}} = 5.31$, $SD_{\text{women}} = 1.35$; $M_{\text{men}} = 4.34$, $SD_{\text{men}} = 1.62$) on scales from 1 (*not at all [trait]*) to 7 (*completely [trait]*). Traits were randomly presented.

Participant Characterization

We averaged responses to the political ideology items (Cronbach's $\alpha = .95$), the hostile sexism items (Cronbach's $\alpha = .92$), the benevolent sexism items (Cronbach's

$\alpha = .88$), the social role belief items (Cronbach's $\alpha = .90$), and the social dominance orientation items. (Cronbach's $\alpha = .91$). See Table 1 for intercorrelations and descriptive statistics.

RESULTS

Gender Differences on the Characterization Measures

Men had more hostile sexism, benevolent sexism, and traditional social role beliefs than women. Men and women were similar in social dominance orientation, political ideology, age, and years of education. See Table 2b.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Trait Evaluations

We regressed evaluations on Job Description ($-1 =$ analytical, $1 =$ interpersonal), Trait (effects coded with competence as the reference level), Facial Characteristics ($-1 =$ masculinized, $1 =$ feminized), Face Gender ($-1 =$ man, $1 =$ woman), Perceiver Gender ($-1 =$ man, $1 =$ woman), and their interactions. Perceiver Gender was included on an exploratory basis. See Table 5 for all coefficient information for the fixed effects. The random effects structure allowed intercepts to vary by participant and face identity and facial characteristics effects to vary by participant and face identity. See Supplemental Materials for main effects and breakdown of each individual trait variable.

Two-Way Interactions

The interaction between Trait and Face Gender replicated Experiment 1 (Figure 2c). Participants evaluated women's (Estimate = 4.49, 95% CI [4.36, 4.63]) versus men's (Estimate = 4.03, 95% CI [3.90, 4.17]) faces as more warm, $b = 0.46$, $SE = 0.08$, $z = 6.02$, $p < .001$. Participants evaluated women's (Estimate = 4.82, 95% CI [4.68, 4.95]) versus men's (Estimate = 4.46, 95% CI [4.32, 4.59]) faces as more competent, $b = 0.36$, $SE = 0.08$, $z = 4.73$, $p < .001$. They similarly evaluated women's (Estimate = 4.42, 95% CI [4.29, 4.56]) and men's (Estimate = 4.51, 95% CI [4.37, 4.65]) dominance, $b = -0.09$, $SE = 0.08$, $z = -1.17$, $p = .09$.

The interaction between Trait and Facial Characteristics partially replicated Experiment 1 (Figure 2d). Like Experiment 1, participants evaluated feminized (Estimate = 4.34, 95% CI [4.23, 4.46]) versus masculinized (Estimate = 4.18, 95% CI [4.06, 4.30]) faces as more warm, $b = 0.16$, $SE = 0.03$, $z = 6.02$, $p < .001$. Also like Experiment 1, they evaluated feminized (Estimate = 4.36, 95% CI [4.25, 4.48]) versus masculinized (Estimate = 4.57, 95% CI [4.45, 4.68]) faces as less dominant, $b = -0.20$, $SE = 0.03$, $z = -7.52$, $p < .001$. Contrasting Experiment 1, they evaluated feminized (Estimate = 4.67, 95% CI [4.55, 4.78]) and masculinized (Estimate = 4.60, 95% CI [4.49, 4.72]) faces as similarly competent, $b = 0.07$, $SE = 0.03$, $z = 2.52$, $p = .12$.

TABLE 5. Effects of Face Gender, Facial Characteristics, Job Description, and Participant Gender on Evaluated Warmth (a), Competence (b), and Dominance (c) in Experiment 2

Predictor	a. Warmth			b. Competence			c. Dominance		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(Intercept)	4.26	[4.11, 4.41]	< .001	4.64	[4.51, 4.76]	< .001	4.47	[4.33, 4.60]	< .001
Face Gender	0.23	[0.12, 0.34]	< .001	0.18	[0.09, 0.27]	< .001	-0.04	[-0.14, 0.05]	.375
Facial Characteristics	0.08	[0.05, 0.11]	< .001	0.03	[0.01, 0.06]	.006	-0.10	[-0.14, -0.07]	< .001
Job Description	0.09	[-0.01, 0.19]	.074	0.10	[0.01, 0.19]	.037	0.07	[-0.02, 0.16]	.112
Participant Gender	-0.02	[-0.12, 0.08]	.693	0.05	[-0.04, 0.15]	.250	0.04	[-0.05, 0.12]	.431
Face Gender × Facial Characteristics	0.03	[-0.00, 0.05]	.071	0.04	[0.01, 0.06]	.003	0.10	[0.07, 0.13]	< .001
Face Gender × Job Description	-0.01	[-0.03, 0.00]	.115	-0.01	[-0.02, 0.01]	.315	0.02	[0.00, 0.04]	.024
Facial Characteristics × Job Description	-0.02	[-0.04, 0.00]	.137	0.00	[-0.01, 0.02]	.646	0.02	[-0.00, 0.04]	.111
Face Gender × Participant Gender	-0.02	[-0.04, -0.01]	.009	0.00	[-0.01, 0.02]	.806	0.01	[-0.01, 0.02]	.454
Facial Characteristics × Participant Gender	-0.01	[-0.04, 0.01]	.144	0.00	[-0.01, 0.02]	.731	0.03	[0.00, 0.05]	.027
Job Description × Participant Gender	0.05	[-0.05, 0.14]	.348	0.02	[-0.07, 0.11]	.660	-0.02	[-0.11, 0.07]	.621
Face Gender × Facial Characteristics × Job Description	-0.01	[-0.02, 0.01]	.537	-0.01	[-0.02, 0.01]	.514	-0.02	[-0.03, 0.00]	.098
Face Gender × Facial Characteristics × Participant Gender	-0.01	[-0.03, 0.01]	.308	0.00	[-0.01, 0.02]	.771	-0.00	[-0.02, 0.01]	.696
Face Gender × Job Description × Participant Gender	0.01	[-0.00, 0.03]	.117	-0.00	[-0.02, 0.01]	.653	-0.01	[-0.02, 0.01]	.441
Facial Characteristics × Job Description × Participant Gender	0.01	[-0.01, 0.03]	.234	-0.00	[-0.02, 0.01]	.629	-0.01	[-0.03, 0.02]	.556
Face Gender × Facial Characteristics × Job Description × Participant Gender	-0.00	[-0.02, 0.01]	.662	-0.00	[-0.02, 0.02]	.978	0.01	[-0.01, 0.03]	.380

Note. Participant Gender: (-1 = men participants, 1 = women participants); Face Gender: (-1 = men's faces, 1 = women's faces); Facial Characteristics: (-1 = masculine, 1 = feminine); Job Description: (-1 = analytical, 1 = interpersonal).

The interaction between Face Gender and Facial Characteristics replicated Experiment 1 (Figure 3b). Participants had higher trait evaluations of feminized (Estimate = 4.63, 95% CI [4.50, 4.77]) versus masculinized (Estimate = 4.52, 95% CI [4.38, 4.65]) women's faces, $b = 0.12$, $SE = 0.03$, $z = 3.93$, $p < .001$. Like Experiment 1, this difference weakened among evaluations of masculinized (Estimate = 4.38, 95% CI [4.24, 4.52]) versus feminized (Estimate = 4.28, 95% CI [4.15, 4.42]) men's faces $b = -0.10$, $SE = 0.03$, $z = -3.32$, $p = .01$.

Three-Way Interactions

Trait, Facial Characteristics, and Face Gender. Trait qualified the interaction between Facial Characteristics and Face Gender (Figure 4b). We characterized this interaction as we did in Experiment 1.

For women, participants evaluated feminized (Estimate = 4.60, 95% CI [4.46, 4.74]) versus masculinized (Estimate = 4.38, 95% CI [4.24, 4.52]) faces as more warm, $b = 0.21$, $SE = 0.04$, $z = 5.79$, $p < .001$. For men, participants similarly evaluated the warmth of feminized (Estimate = 4.09, 95% CI [3.95, 4.23]) and masculinized (Estimate = 3.98, 95% CI [3.84, 4.12]) faces $b = 0.11$, $SE = 0.04$, $z = 2.94$, $p = .13$.

For women, participants evaluated feminized (Estimate = 4.88, 95% CI [4.74, 5.02]) versus masculinized (Estimate = 4.75, 95% CI [4.61, 4.89]) faces as more competent, $b = 0.14$, $SE = 0.04$, $z = 3.75$, $p = .01$. For men, participants similarly evaluated the competency of feminized (Estimate = 4.45, 95% CI [4.31, 4.59]) versus masculinized (Estimate = 4.46, 95% CI [4.32, 4.60]) faces, $b = -0.003$, $SE = 0.04$, $z = -0.09$, $p > .99$.

For women, participants similarly evaluated the dominance of feminized (Estimate = 4.42, 95% CI [4.28, 4.56]) versus masculinized (Estimate = 4.42, 95% CI [4.28, 4.56]) faces, $b = -0.003$, $SE = 0.04$, $z = -0.09$, $p > .99$. For men, participants evaluated masculinized (Estimate = 4.71, 95% CI [4.57, 4.85]) versus feminized (Estimate = 4.31, 95% CI [4.17, 4.45]) faces as more dominant, $b = -0.40$, $SE = 0.04$, $z = -10.91$, $p < .001$.

Trait, Face Gender, and Perceiver Gender. Trait qualified the interaction between Face Gender and Perceiver Gender (Figure 5b). We characterized this interaction as we did in Experiment 1.

Men evaluated men's faces as similarly dominant (Estimate = 4.48, 95% CI [4.33, 4.64]) and competent (Estimate = 4.40, 95% CI [4.25, 4.56]), $b = -0.13$, $SE = 0.04$, $z = -3.515$, $p = .02$. In contrast, they evaluated women's faces as more competent (Estimate = 4.76, 95% CI [4.60, 4.91]) than dominant (Estimate = 4.38, 95% CI [4.22, 4.53]), $b = 0.54$, $SE = 0.04$, $z = 14.49$, $p < .001$. Men evaluated men's faces as more competent (Estimate = 4.40, 95% CI [4.25, 4.56]) than warm (Estimate = 4.03, 95% CI [3.87, 4.18]), $b = 0.38$, $SE = 0.04$, $z = 10.36$, $p < .001$. They also rated women's faces as more competent (Estimate = 4.76, 95% CI [4.60, 4.91]) than warm (Estimate = 4.53, 95% CI [4.38, 4.69]), $b = 0.28$, $SE = 0.04$, $z = 7.63$, $p < .001$. Men evaluated men's faces as more dominant (Estimate = 4.48, 95% CI [4.33, 4.64]) than

warm (Estimate = 4.03, 95% CI [3.87, 4.18]), $b = 0.51$, $SE = 0.04$, $z = 13.88$, $p < .001$. In contrast, they rated women's faces as more warm (Estimate = 4.53, 95% CI [4.38, 4.69]) than dominant (Estimate = 4.38, 95% CI [4.22, 4.53]), $b = -0.25$, $SE = 0.04$, $z = -6.86$, $p < .001$.

Like men, women evaluated men's faces as dominant (Estimate = 4.54, 95% CI [4.37, 4.71]) and competent (Estimate = 4.51, 95% CI [4.34, 4.67]), $b = -0.03$, $SE = 0.03$, $z = -1.13$, $p = .99$. Also like men, women evaluated women's faces as more competent (Estimate = 4.87, 95% CI [4.71, 5.04]) than dominant (Estimate = 4.46, 95% CI [4.30, 4.63]), $b = 0.41$, $SE = 0.03$, $z = 14.56$, $p < .001$. Like men, women evaluated men's faces as more competent (Estimate = 4.51, 95% CI [4.34, 4.67]) than warm (Estimate = 4.03, 95% CI [3.87, 4.20]), $b = 0.47$, $SE = 0.03$, $z = 16.80$, $p < .001$. Also like men, they rated women's faces as higher in competence (Estimate = 4.87, 95% CI [4.71, 5.04]) than warmth (Estimate = 4.45, 95% CI [4.28, 4.62]), $b = 0.42$, $SE = 0.03$, $z = 15.09$, $p < .001$. Like men, women evaluated men's faces as more dominant (Estimate = 4.54, 95% CI [4.37, 4.71]) than warm (Estimate = 4.03, 95% CI [3.87, 4.20]), $b = 0.50$, $SE = 0.03$, $z = 17.93$, $p < .001$. In contrast to men and replicating Experiment 1, women rated women's faces as similarly warm (Estimate = 4.45, 95% CI [4.28, 4.62]) and dominant (Estimate = 4.46, 95% CI [4.30, 4.63]), $b = 0.01$, $SE = 0.03$, $z = 0.53$, $p > .99$.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

We again created a composite outcome rating by averaging across the three outcomes (Cronbach's $\alpha = .93$). We regressed these composite ratings on Job Description ($-1 = \text{analytical}$, $1 = \text{interpersonal}$), Facial Characteristics ($-1 = \text{masculinized}$, $1 = \text{feminized}$), Face Gender ($-1 = \text{man}$, $1 = \text{woman}$), Perceiver Gender ($-1 = \text{man}$, $1 = \text{woman}$), and their interactions as fixed effects. Perceiver Gender was included on an exploratory basis. See Table 6 for all model coefficients for the fixed effects. The model included a random effects structure that allowed intercepts to vary by participant and face identity and allowed Facial Characteristics effects to vary by participant and face identity. See Supplemental Materials for main effects and breakdown of each individual outcome variable.

Two-Way Interaction

Replicating Experiment 1, Face Gender and Facial Characteristics interacted (Figure 6b). Participants had higher endorsements of feminized (Estimate = 4.50, 95% CI [4.33, 4.68]) versus masculinized (Estimate = 4.32, 95% CI [4.14, 4.50]) women's faces, $b = 0.18$, $SE = 0.04$, $z = 4.34$, $p < .001$. Participants similarly endorsed feminized (Estimate = 4.02, 95% CI [3.84, 4.19]) and masculinized (Estimate = 4.02, 95% CI [3.84, 4.20]) men's faces, $b = -0.01$, $SE = 0.04$, $z = -0.05$, $p = .99$. A sensitivity analysis revealed that, given our specified model and sample size, we could detect an interactive effect size of 0.026 with 80% power and $\alpha = .05$.

TABLE 6. Effects of Face Gender, Facial Characteristics, Job Description, and Participant Gender on (a) Candidate Endorsement, (b) Likelihood of Doing Well in Position, and (c) People Liking Candidate as a Leader in Experiment 2

Predictor	a. Candidate Endorsement			b. Likelihood of Doing Well in Position			c. People Liking Candidate as a Leader		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(Intercept)	4.14	[4.00, 4.29]	< .001	4.29	[4.15, 4.42]	< .001	4.22	[4.08, 4.36]	< .001
Face Gender	0.21	[0.11, 0.32]	< .001	0.20	[0.10, 0.30]	< .001	0.18	[0.08, 0.28]	.001
Facial Characteristics	0.04	[0.01, 0.08]	.010	0.04	[0.01, 0.07]	.009	0.05	[0.02, 0.09]	.001
Job Description	0.12	[0.02, 0.22]	.023	0.07	[−0.02, 0.17]	.129	0.10	[0.00, 0.20]	.041
Participant Gender	−0.02	[−0.13, 0.08]	.656	−0.04	[−0.13, 0.06]	.456	−0.02	[−0.12, 0.07]	.622
Face Gender × Facial Characteristics	0.04	[0.01, 0.08]	.005	0.04	[0.01, 0.07]	.003	0.05	[0.02, 0.08]	< .001
Face Gender × Job Description	0.01	[−0.01, 0.03]	.395	0.01	[−0.01, 0.03]	.357	0.00	[−0.01, 0.02]	.680
Facial Characteristics × Job Description	0.01	[−0.01, 0.03]	.354	0.01	[−0.01, 0.03]	.489	0.00	[−0.02, 0.02]	.946
Face Gender × Participant Gender	0.01	[−0.01, 0.03]	.431	0.01	[−0.01, 0.02]	.449	−0.02	[−0.04, −0.00]	.040
Facial Characteristics × Participant Gender	−0.01	[−0.03, 0.01]	.316	−0.02	[−0.04, 0.01]	.141	−0.01	[−0.03, 0.02]	.497
Job Description × Participant Gender	0.01	[−0.09, 0.11]	.830	0.01	[−0.08, 0.11]	.765	−0.01	[−0.11, 0.09]	.863
Face Gender × Facial Characteristics × Job Description	−0.02	[−0.04, −0.00]	.029	−0.01	[−0.03, 0.00]	.109	−0.01	[−0.03, 0.01]	.343
Face Gender × Facial Characteristics × Participant Gender	0.00	[−0.02, 0.02]	.790	−0.00	[−0.02, 0.02]	.887	−0.00	[−0.02, 0.02]	.847
Face Gender × Job Description × Participant Gender	−0.00	[−0.02, 0.02]	.888	−0.01	[−0.02, 0.01]	.449	0.00	[−0.02, 0.02]	.812
Facial Characteristics × Job Description × Participant Gender	−0.01	[−0.04, 0.01]	.171	−0.01	[−0.03, 0.01]	.521	−0.00	[−0.03, 0.02]	.793
Face Gender × Facial Characteristics × Job Description × Participant Gender	−0.01	[−0.02, 0.01]	.523	−0.01	[−0.02, 0.01]	.487	−0.01	[−0.03, 0.00]	.144

Note. Participant Gender: (−1 = men participants, 1 = women participants); Face Gender: (−1 = women's faces, 1 = men's faces); Facial Characteristics: (−1 = analytical, 1 = interpersonal).
Description: (−1 = analytical, 1 = interpersonal).

Linking Beliefs About College Students to Outcomes

We next examined whether beliefs about college students related to women versus men being more favorably evaluated. Because we wanted to explore whether *gender-biased beliefs* about college students' traits qualified Face Gender effects, we created difference scores for beliefs about warmth, competence, dominance, and responsibility by subtracting the belief about men from the belief about women. Higher scores reflect more attribution to women than men. We standardized difference scores around the mean for each trait. Participants believed that women were warmer, $t(372) = 11.01, p < .002, d = 0.77$, more competent, $t(372) = 6.11, p < .001, d = 0.42$, and more responsible, $t(372) = 12.26, p < .001, d = 0.97$, than men. They believed men were more dominant than women, $t(372) = -11.50, p < .001, d = 0.92$.

We regressed the composite outcome rating on Face Gender ($-1 = \text{man}, 1 = \text{woman}$), standardized beliefs about students' Warmth, Competence, Dominance, and Responsibility, as well as interactive effects between Face Gender and each belief. See Table 7 for coefficient information for these fixed effects. The random effects structure allowed intercepts to vary by participant and face identity.

Face Gender and Warmth beliefs interacted (Figure 7a). Believing that college women are warmer than men negatively related to outcomes for men, $b = -0.24, SE = 0.05, z = -4.32, p < .0001$. To a lesser extent, it negatively related to outcomes for women, $b = -0.12, SE = 0.05, z = -2.17, p = .03$. Face Gender and Competence beliefs interacted (Figure 7b). Believing that college women are more competent than men positively related to outcomes for women, $b = 0.16, SE = 0.06, z = 2.78, p = .005$. It did not significantly relate to outcomes for men, $b = 0.01, SE = 0.06, z = 0.21, p = .83$. Face Gender and Responsibility beliefs interacted (Figure 7c). Believing that college women are more responsible than men negatively related to outcomes for men, $b = -0.21, SE = 0.06, z = -3.65, p < .001$. It did not significantly relate to outcomes for women, $b = -0.09, SE = 0.06, z = -1.58, p = .11$.

DISCUSSION

Gender-biased beliefs about college-aged women and men related to their endorsements of men and women for the student leadership position. These patterns complement work showing that gender-related attitudes affect evaluations of aspiring leaders (e.g., Carlin & Winfrey, 2009). The extent to which people believed college-aged women versus men were warmer and more responsible more strongly negatively related to outcomes for men than women. To the extent that responsible people are trustworthy and thus high in warmth (Fiske et al., 2007), believing that college-aged women versus men are warmer negatively affects how people evaluate men. These traits reflect a broader valence dimension (Oosterhof & Todorov, 2008). This pattern may thus reflect that perceiving college-aged men more negatively than women negatively affects how people evaluate men for leadership.

The extent to which people believed college-aged women versus men were more competent more positively related to outcomes for women versus men. That women were believed to be more competent than men reflects growing societal beliefs in competence equality between men and women (Eagly et al., 2020) and

TABLE 7. Effects of Face Gender, Warmth-Related Gender Stereotype, Competence-Related Gender Stereotype, Dominance-Related Gender Stereotype, and Responsibility-Related Gender Stereotype on All Evaluated Outcomes (Combined) in Experiment 2

Predictor	Evaluated Outcome		
	Estimate	CI	<i>p</i>
(Intercept)	4.21	[4.08, 4.35]	< .001
Face Gender	0.20	[0.10, 0.30]	< .001
Warmth-related gender stereotype	−0.18	[−0.28, −0.07]	.001
Competence-related gender stereotype	0.08	[−0.02, 0.19]	.128
Dominance-related gender stereotype	0.06	[−0.03, 0.16]	.203
Responsibility-related gender stereotype	−0.15	[−0.26, −0.04]	.008
Face Gender × Warmth-related gender stereotype	0.06	[0.04, 0.08]	< .001
Face Gender × Competence-related gender stereotype	0.07	[0.05, 0.09]	< .001
Face Gender × Dominance-related gender stereotype	−0.00	[−0.02, 0.02]	.790
Face Gender × Responsibility-related gender stereotype	0.06	[0.04, 0.08]	< .001

may tip toward gender-biased beliefs in some populations. That gender-biased beliefs favorably related to evaluations of women also suggests that women do not always incur a penalty for being high in an agentic quality (i.e., competence). In fact, they may sometimes be rewarded for it. Future work may test if competency relates to favorable outcomes for women in the absence of favorable gender-biased beliefs relative to other agentic traits. Women may incur negative bias in contexts where beliefs misalign with their success. One possibility is that because beliefs about college students were assessed after the main task, the main task could have influenced these beliefs. Although we cannot be certain about this possibility, the job descriptions did not elicit any differences in gender-biased beliefs (see Supplemental Materials). At the very least, our between-participants manipulation did not unduly influence beliefs. Because attitudes are malleable based on context (Blair, 2002), future work should directly test this possibility by manipulating task order. Furthermore, because our link from beliefs to outcomes is correlational, future work can build on these findings by manipulating beliefs to determine causality.

The data did not support women being more favorably evaluated only for a more feminine role. If anything, it ruled against this possibility. Indeed, whereas endorsing men and women candidates for the analytical role differed by sexually dimorphic facial characteristics as in Experiment 1, this pattern weakened for endorsements within the interpersonal role. This finding counters the belief that people support women leaders more in feminine than in neutral and masculine contexts (e.g., Banchevsky et al., 2016; Giacomini et al., 2022). It may be that facial versus behavioral cues do not as strongly affect evaluations by job context.

Beyond examining the overall target gender effect on outcomes, Experiment 2 served as a replication of Experiment 1. The findings of Experiment 1 were largely

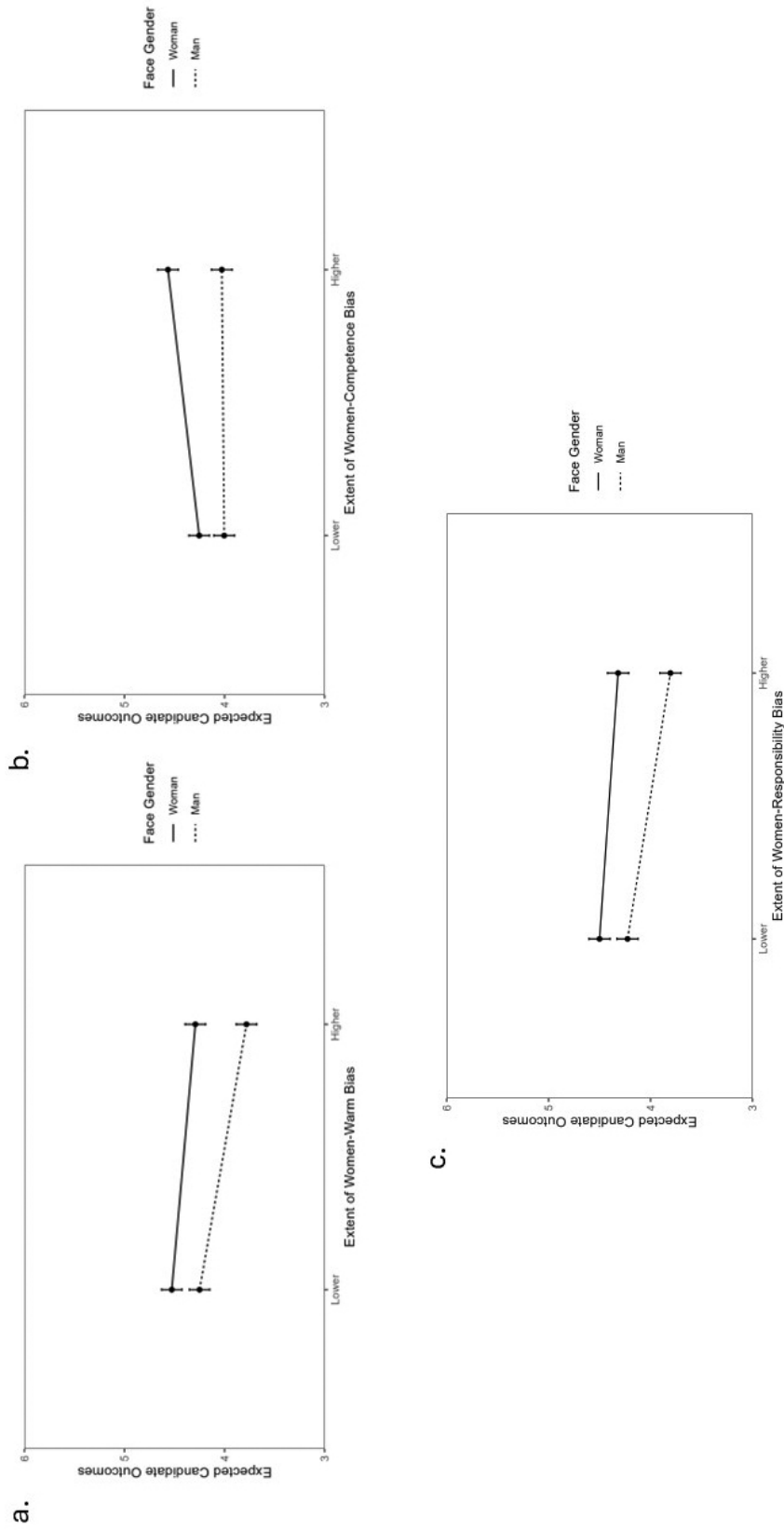


FIGURE 7. In Experiment 2, Face Gender by Trait-related warmth (a), competence (b), and responsibility (c) gender stereotypes interactions emerged for candidate evaluation positivity.

replicated in Experiment 2 using a new and larger participant sample. Replicating Experiment 1 were trait attributions based on gender and sexually dimorphic facial features. Also replicating Experiment 1, differential evaluations only emerged for women based on sexually dimorphic facial features. Furthermore, feminized women's faces again received the highest endorsements relative to all other combinations of facial characteristics and target gender (see Supplemental Materials).

GENERAL DISCUSSION

Sexually dimorphic facial features were supported as a barrier to aspiring women leaders receiving support. Aligning with expectations from backlash research (Williams & Tiedens, 2016), women with masculinized versus feminized features received more negative evaluations in terms of their leadership suitability, but evaluations of men were not differentially affected. Contrasting these expectations, however, women and feminized faces received consistently more favorable evaluations. In fact, feminized women's faces received the most favorable evaluative outcomes relative to the other combinations of target gender and facial characteristics across experiments. Rather than showing that femininity exclusively affects female leaders negatively (e.g., Banchevsky et al., 2016), we show that, at least in terms of nonverbal cues, femininity can be a benefit for aspiring White women leaders in some contexts. These patterns were robust to roles varying in their interpersonal or analytical nature. By having participants evaluate *aspiring* rather than *established* (see Joshi et al., 2023) women leaders, this work informs how White women are *selected* to leadership roles on the basis of sexually dimorphic facial features.

Prior work posits that aspiring women leaders are perceived as agentic (Bosak & Sczesny, 2011; Rudman & Glick, 1999). Role prioritization theory, however, posits that aspiring women leaders perceived as agentic *and* communal receive more positive evaluations than women perceived as one or the other (Haines & Stroessner, 2019). Feminized features may provide a pathway to perceiving the simultaneous agency and communality that specifically benefits women (Gervais & Hillard, 2011). Indeed, although people endorse Kamala Harris as being highly agentic (Liebenow et al., 2022), people who have a feminized representation of her face most favorably evaluated her 2020 election candidacy (Cassidy & Liebenow, 2021). Extending the literature, the current work supports the possibility that these simultaneously perceived competence and dominance inferences with warmth inferences can be acquired from different sources and integrated in evaluations of women. Such a possibility reflects a dynamic impression formation process in which cues from different sources are integrated in a coherent percept (Freeman & Ambady, 2011).

That women with masculinized versus feminized faces received more negative outcomes could be construed as nonverbal backlash. Yet women were consistently favored over men for the student leadership role. Beliefs that college-aged women versus men were more warm, competent, and responsible positively related to these favorable outcomes for women. Notably, feminized women's faces received the most positive outcomes relative to all other combinations of target facial features and genders. Benefits for women, and especially women with feminized

faces, contrast with a backlash pattern. At first blush, this possibility may seem positive for women. Yet manipulating sexually dimorphic facial features affected only women candidates. Thus, even when women might “have it all,” so to speak, they can still be penalized for possessing stereotypically masculine qualities. We suggest that sexually dimorphic facial features comprise a nonverbal barrier specific to aspiring women leaders.

That feminized women’s faces received the most positive outcomes raises the possibility that effects of sexually dimorphic facial features reflect a halo from which feminized women benefit rather than a backlash against masculinized women. Notably, college-aged women versus men were believed to be more warm, competent, and responsible. These traits convey positive valence. Although men were believed to be more dominant than women, dominance is separable from valence in terms of evaluations of others (Oosterhof & Todorov, 2008). Speculatively, because college-aged women are more positively evaluated than men, they might also be more *attractive* to people making social decisions. Thus, the overall favoritism received by women could reflect this attractiveness.

Supporting this possibility, attractiveness confers stronger inferences of sophisticated minds (Sherman & Haidt, 2011), mental stability (Farina et al., 1977), and social value (Fink & Penton-Voak, 2002). Attractiveness also more strongly affects evaluations of women’s relative to men’s mental sophistication (Alaei et al., 2022). Because feminized versus masculinized faces are evaluated as being more attractive (Perrett et al., 1998), it could be that a halo inherent to attitudes about college-aged women is accentuated among women with more feminized facial features. This halo may also reflect a preference for more stereotypically and conventionally “attractive” women reflected in their leadership prospects (Hoss et al., 2005; Hu et al., 2018; Perrett et al., 1998) versus a halo reflecting that these women are more mentally able. Future work may disentangle these possibilities. Such halos, however, are likely context-dependent. Indeed, women with feminized faces are often not considered suitable for agentic leadership roles (Banchefsky et al., 2016; Joshi et al., 2023). Future work could collect attractiveness ratings of targets and manipulate the context of the leadership role to examine a unique role for attractiveness and a potential attractiveness halo in outcomes for aspiring women leaders. If such a halo exists, it would reflect a nonverbal route contributing to disparities in leadership among women potentially separable from the backlash shown in prior work (e.g., Rudman & Glick, 1999).

In the current work, participants viewed candidate faces and likely made agency-related trait inferences about all targets via their interest in the leadership position. Although women were evaluatively penalized for having masculinized versus feminized facial features, an open question regards how much feminized facial features might help these women once they gain leadership positions. Such features may help agentic women to obtain leadership (Cassidy & Liebenow, 2021), but people may consistently expect these women’s communality even when they enact agentic behaviors. Speculatively, feminized features may be helpful for women to obtain but not *maintain* leadership. Indeed, although face impressions are enduring, they are not impenetrable to updating based on incoming behavioral

information (Shen & Ferguson, 2021). Future work may examine the extent to which agentic women with feminized facial features are evaluated as more agentic or less communal as they enact leadership-related behaviors.

Because the current work used only White faces, it does not address intersectionality's likely contributions to evaluations of aspiring women leaders. These contributions reflect an important next step for future work. People are slower, for example, to categorize Black versus Asian women as being women (Johnson et al., 2012). Furthermore, racial stereotypes elicit a tendency to overlook Black women's femininity (Takinami, 2016). Building on these findings, one possibility is that feminized versus masculinized Black women's faces may not receive as favorable outcomes in terms of leadership endorsements as White women's faces do. This possibility is critical to assess given that inferences of communal traits like warmth are core to Black individuals being endorsed for their leadership abilities (Wilson et al., 2017). It is also important to not only consider one category at a time because people identify with more than one. Supporting the importance of this future direction, people broadly perceive gay Black men as better leaders than straight Black and gay White men, in part due to perceptions of masculinity and warmth (Wilson et al., 2017). Gender may further complicate these patterns in ways that are important to disentangle given calls for understanding leadership in the context of diversity (Eagly & Chin, 2010).

Another potential limitation here is that agency- and communion-related inferences were constrained to competence, dominance, and warmth. Because *agency* and *communion* are likely not familiar terms outside of academic discourse, we used traits more familiar to a general population and that have been used in related work (Abele et al., 2016; Fiske et al., 2002, 2007; Imhoff et al., 2013; Wen et al., 2020). Future work may benefit from using different traits related to agency and communality to assess the generalizability of the current findings. For example, whereas warmth is an oft-tested communal trait, it can be argued that it is only a singular component of the larger concept of communion. Thus, future work may benefit from assessing different communal traits (e.g., caring) and how their inferences are differentially elicited by sexually dimorphic facial features. It could be that some communal traits are more strongly inferred by these features than others.

Here, we identify relative sexually dimorphic facial features as a potential barrier that young women aspiring to leadership may face early on in their careers. These patterns call for work focusing on when (e.g., stage of life) and where (e.g., workplace environment) women's appearance may limit their leadership prospects. Importantly, we do not endorse a quick fix that women should "just be more feminine" to get ahead. Rather, we view these findings as a call for more research aimed at reducing pervasive facial appearance effects on outcomes (e.g., Chua & Freeman, 2021). Such research can inform targeted interventions to support women's leadership development at different life stages, such as training for hiring committees or considerations when creating media and marketing campaigns. This work may be particularly important because having especially feminine role models in STEM actually demotivates girls from pursuing STEM because this combination of traits seems unattainable (Betz & Sekaquaptewa, 2012). If women

see only very feminized women achieving leadership roles, it could demotivate them from pursuing leadership and perpetuate gender disparity in leadership.

Examining contributions of sexually dimorphic facial features to outcomes for aspiring women leaders enhances our understanding of gender disparities in leadership. Having more feminized versus masculinized sexually dimorphic features benefits college-aged women regardless of whether a job is more stereotypically feminine versus masculine. Future work may build upon the provided foundation to form a more comprehensive picture of the challenges that aspiring leaders face through nonverbal cues they may not be able to control.

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>
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, 93(5), 751–763. <https://doi.org/10.1037/0022-3514.93.5.751>
- Alaei, R., Deska, J. C., Hugenberg, K., & Rule, N. O. (2022). People attribute humanness to men and women differently based on their facial appearance. *Journal of Personality and Social Psychology*, 123, 400–422. <https://doi.org/10.1037/pspi0000364>
- Baber, K. M., & Tucker, C. J. (2006). The Social Roles Questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles: A Journal of Research*, 54(7–8), 459–467. <https://doi.org/10.1007/s11199-006-9018-y>
- Banchefsky, S., Westfall, J., Park, B., & Judd, C. M. (2016). But you don't look like a scientist!: Women scientists with feminine appearance are deemed less likely to be scientists. *Sex Roles*, 75(3–4), 95–109. <https://doi.org/10.1007/s11199-016-0586-1>
- Baskaran, T., & Hessami, Z. (2018). Does the election of a female leader clear the way for more women in politics? *American Economic Journal: Economic Policy*, 10(3), 95–121. <https://doi.org/10.1257/pol.20170045>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauman, K., & Cranney, S. (2020). School enrollment in the United States: 2018. Population characteristics. *Current Population Reports* (pp. 20–584). U.S. Census Bureau. <https://www.census.gov/content/dam/Census/library/publications/2020/demo/p20-584.pdf>
- Belanger, A. L., Joshi, M. P., Fuesting, M. A., Weisgram, E. S., Claypool, H. M., & Diekman, A. B. (2020). Putting belonging in context: Communal affordances signal belonging in STEM. *Personality and Social Psychology Bulletin*, 46(8), 1186–1204. <https://doi.org/10.1177/0146167219897181>
- Betz, D., & Sekaquaptewa, D. (2012). My fair physicist? Feminine math and science role models demotivate young girls. *Social Psychological and Personality Science*, 3(6), 738–746. <https://doi.org/10.1177/1948550612440735>
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8
- 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(3), 560–597. <https://doi.org/10.1111/j.1559-1816.2011.00781.x>
- Bosak, J., & Sczesny, S. (2011). Exploring the dynamics of incongruent beliefs about women and leaders. *British Journal of Management*, 22(2), 254–269. <https://doi.org/10.1111/j.1467-8551.2010.00731.x>

- Broockman, D. E. (2014). Do female politicians empower women to vote or run for office? A regression discontinuity approach. *Electoral Studies*, 34, 190–204. <https://doi.org/10.1016/j.electstud.2013.10.002>
- Bruckmüller, S., & Abele, A. E. (2013). The density of the big two: How are agency and communion structurally represented? *Social Psychology*, 44(2), 63–74. <https://doi.org/10.1027/1864-9335/a000145>
- Carlin, D. B., & Winfrey, K. L. (2009). Have you come a long way, baby? Hillary Clinton, Sarah Palin, and sexism in 2008 campaign coverage. *Communication Studies*, 60(4), 326–343. <https://doi.org/10.1080/10510970903109904>
- Carpinella, C. M., & Johnson, K. L. (2016). Visual political communication: The impact of facial cues from social constituencies to personal pocketbooks. *Social and Personality Psychology Compass*, 10(5), 281–297. <https://doi.org/10.1111/spc3.12249>
- Cassidy, B. S., & Liebenow, H. A. (2021). Feminine perceptions of Kamala Harris positively relate to evaluations of her candidacy. *Analyses of Social Issues and Public Policy (ASAP)*, 21(1), 29–50. <https://doi.org/10.1111/asap.12243>
- Chua, K.-W., & Freeman, J. B. (2021). Facial stereotype bias is mitigated by training. *Social Psychological and Personality Science*, 12(7), 1335–1344. <https://doi.org/10.1177/1948550620972550>
- Conroy, M., Joesten Martin, D., & Nalder, K. L. (2020). Gender, sex, and the role of stereotypes in evaluations of Hillary Clinton and the 2016 presidential candidates. *Journal of Women, Politics & Policy*, 41(2), 194–218. <https://doi.org/10.1080/1554477X.2020.1731280>
- Creamer, E. G. (2012). Effects of numeric representation of women on interest in engineering as a career. *American Journal of Engineering Education (AJEE)*, 3(1), 1–12. <https://doi.org/10.19030/ajee.v3i1.6884>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40, 61–149. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- DeBruine, L., & Jones, B. (2017). *Face Research Lab London Set (Version 3)* (p. 281699312 Bytes) [dataset]. [object Object]. <https://doi.org/10.6084/M9.FIGSHARE.5047666.V3>
- Ditonto, T., & Mattes, K. (2018). Differences in appearance-based trait inferences for male and female political candidates. *Journal of Women, Politics & Policy*, 39(4), 430–450. <https://doi.org/10.1080/1554477X.2018.1506206>
- Eagly, A. H., & Chin, J. L. (2010). Diversity and leadership in a changing world. *American Psychologist*, 65(3), 216–224. <https://doi.org/10.1037/a0018957>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, 75(3), 301–315. <https://doi.org/10.1037/amp0000494>
- Farina, A., Fischer, E. H., Sherman, S., Smith, W. T., Groh, T., & Mermin, P. (1977). Physical attractiveness and mental illness. *Journal of Abnormal Psychology*, 86(5), 510–517. <https://doi.org/10.1037/0021-843X.86.5.510>
- Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science*, 11(5), 154–158. <https://doi.org/10.1111/1467-8721.00190>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82, 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>

- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, 118(2), 247–279. <https://doi.org/10.1037/a0022327>
- Gadassi, R., & Gati, I. (2009). The effect of gender stereotypes on explicit and implicit career preferences. *Counseling Psychologist*, 37(6), 902–922. <https://doi.org/10.1177/0011000009334093>
- Gervais, S. J., & Hillard, A. L. (2011). A role congruity perspective on prejudice toward Hillary Clinton and Sarah Palin. *Analyses of Social Issues and Public Policy*, 11(1), 221–240. <https://doi.org/10.1111/j.1530-2415.2011.01263.x>
- Giacomin, M., Tskhay, K. O., & Rule, N. O. (2022). Gender stereotypes explain different mental prototypes of male and female leaders. *Leadership Quarterly*, 33(6), 101578. <https://doi.org/10.1016/j.leaqua.2021.101578>
- 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., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles*, 32(9–10), 565–582. <https://doi.org/10.1007/BF01544212>
- Goh, J. X., Bandt-Law, B., Cheek, N. N., Sinclair, S., & Kaiser, C. R. (2022). Narrow prototypes and neglected victims: Understanding perceptions of sexual harassment. *Journal of Personality and Social Psychology*, 122(5), 873–893. <https://doi.org/10.1037/pspi0000260>
- Goodale, B. M., Alt, N. P., Lick, D. J., & Johnson, K. L. (2018). Groups at a glance: Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *Journal of Experimental Psychology: General*, 147(11), 1660–1676. <https://doi.org/10.1037/xge0000450>
- Gundersen, A. B., & Kunst, J. R. (2019). Feminist ≠ Feminine? Feminist women are visually masculinized whereas feminist men are feminized. *Sex Roles*, 80(5), 291–309. <https://doi.org/10.1007/s11199-018-0931-7>
- Haines, E. L., & Stroessner, S. J. (2019). The role prioritization model: How communal men and agentic women can (sometimes) have it all. *Social and Personality Psychology Compass*, 13(12), e12504. <https://doi.org/10.1111/spc3.12504>
- Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science*, 5(7), 815–824. <https://doi.org/10.1177/1948550614534701>
- Hester, N., Jones, B. C., & Hehman, E. (2021). Perceived femininity and masculinity contribute independently to facial impressions. *Journal of Experimental Psychology: General*, 150(6), 1147–1164. <https://doi.org/10.1037/xge0000989>
- Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO_x scale. *Journal of Personality and Social Psychology*, 109(6), 1003–1028. <https://doi.org/10.1037/pspi0000033>
- Hoss, R. A., Ramsey, J. L., Griffin, A. M., & Langlois, J. H. (2005). The role of facial attractiveness and facial masculinity/femininity in sex classification of faces. *Perception*, 34(12), 1459–1474. <https://doi.org/10.1068/p5154>
- Hu, Y., Abbasi, N. U. H., Zhang, Y., & Chen, H. (2018). The effect of target sex, sexual dimorphism, and facial attractiveness on perceptions of target attractiveness and trustworthiness. *Frontiers in Psychology*, 9, 942. <https://doi.org/10.3389/fpsyg.2018.00942>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, 4, 386. <https://doi.org/10.3389/fpsyg.2013.00386>
- Jackson, L. A., Sullivan, L. A., & Hymes, J. S. (1987). Gender, gender role, and physical appearance. *Journal of Psychology*, 121(1), 51–56. <https://doi.org/10.1080/00223980.1987.9712642>
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality*

- and *Social Psychology*, 102(1), 116–131. <https://doi.org/10.1037/a0025335>
- Joshi, M. P., Lloyd, E. P., Diekman, A. B., & Hugenberg, K. (2023). In the face of opportunities: Facial structures of scientists shape expectations of STEM environments. *Personality and Social Psychology Bulletin*, 49(5), 673–691. <https://doi.org/10.1177/01461672221077801>
- Kahn, S., & Ginther, D. K. (2017). *Women and STEM* (SSRN Scholarly Paper ID 2988746). Social Science Research Network. <https://papers.ssrn.com/abstract=2988746>
- Kleider-Offutt, H., Meacham, A. M., Brnum-Martin, L., & Capodanno, M. (2021). What's in a face? The role of facial features in ratings of dominance, threat, and stereotypicality. *Cognitive Research: Principles and Implications*, 6(1), 53. <https://doi.org/10.1186/s41235-021-00319-9>
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. <https://doi.org/10.1037/a0023557>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package “Emmeans”. R Package Version 4.0-3. <http://cran.r-project.org/package=emmeans>
- Lick, D. J., & Johnson, K. L. (2014). Recalibrating gender perception: Face aftereffects and the perceptual underpinnings of gender-related biases. *Journal of Experimental Psychology: General*, 143(3), 1259–1276. <https://doi.org/10.1037/a0034516>
- Liebenow, H. A., Boucher, K. L., & Cassidy, B. S. (2022). Understanding evaluations of Kamala Harris in 2020: Political ideology qualifies perceived communality effects when communal cues are present. *Psychology of Women Quarterly*, 46(4), 399–419. <https://doi.org/10.1177/03616843221104383>
- Little, A. C. (2014). Facial attractiveness. *WIREs Cognitive Science*, 5(6), 621–634. <https://doi.org/10.1002/wcs.1316>
- Little, A., Jones, B., & DeBruine, L. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366, 1638–1659. <https://doi.org/10.1098/rstb.2010.0404>
- Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W. Y., Kidd, C., Marcelle, E., . . . Dronkers, N. F. (2021). Gender bias in academia: A life-time problem that needs solutions. *Neuron*, 109(13), 2047–2074. <https://doi.org/10.1016/j.neuron.2021.06.002>
- Marcinkowska, U. M., Kozlov, M. V., Cai, H., Contreras-Garduño, J., Dixon, B. J., Oana, G. A., Kaminski, G., Li, N. P., Lyons, M. T., Onyishi, I. E., Prasai, K., Pazhoohi, F., Prokop, P., Rosales Cardozo, S. L., Sydney, N., Yong, J. C., & Rantala, M. J. (2014). Cross-cultural variation in men's preference for sexual dimorphism in women's faces. *Biology Letters*, 10(4), 20130850. <https://doi.org/10.1098/rsbl.2013.0850>
- Martin, A. E., & Slepian, M. L. (2020). The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. *Perspectives on Psychological Science*, 174569162090496. <https://doi.org/10.1177/1745691620904961>
- Moss-Racusin, C. A. (2014). Male backlash: Penalties for men who violate gender stereotypes. In R. J. Burke & D. A. Major (Eds.), *Gender in organizations: Are men allies or adversaries to women's career advancement?* (pp. 247–269). Edward Elgar Publishing. <https://doi.org/10.4337/9781781955703.00021>
- Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, 11(2), 140–151. <https://doi.org/10.1037/a0018093>
- Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science*, 30(1), 65–79. <https://doi.org/10.1177/0956797618813092>
- Oh, D., Dotsch, R., Porter, J., & Todorov, A. (n.d.). *Gender biases in impressions from faces: Empirical studies and computational models*.

- Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, 36(7), 923–936. <https://doi.org/10.1177/0146167210371949>
- Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2020). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two? *European Journal of Social Psychology*, 50(2), 463–483. <https://doi.org/10.1002/ejsp.2618>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L., & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, 394(6696), Article 6696. <https://doi.org/10.1038/29772>
- 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>
- Piatek-Jimenez, K., Cribbs, J., & Gill, N. (2018). College students' perceptions of gender stereotypes: Making connections to the underrepresentation of women in STEM fields. *International Journal of Science Education*, 40(12), 1432–1454. <https://doi.org/10.1080/09500693.2018.1482027>
- 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>
- Quittkat, H. L., Hartmann, A. S., Düsing, R., Buhlmann, U., & Vocks, S. (2019). Body dissatisfaction, importance of appearance, and body appreciation in men and women over the lifespan. *Frontiers in Psychiatry*, 10, 864. <https://doi.org/10.3389/fpsy.2019.00864>
- Re, D. E., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2013). Facial cues to perceived height influence leadership choices in simulated war and peace contexts. *Evolutionary Psychology*, 11(1), 89–103.
- Re, D. E., & Rule, N. (2017). Distinctive Facial Cues Predict Leadership Rank and Selection. *Personality & social psychology bulletin*, 43(9), 1311–1322. <https://doi.org/10.1177/0146167217712989>
- Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(Suppl. 1), S93–S95. <https://doi.org/10.1098/rsbl.2003.0023>
- Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: how role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, 95(2), 221–235. <https://doi.org/10.1037/a0018204>
- 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), 629–645. <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(5), 1004–1010. <https://doi.org/10.1037/0022-3514.77.5.1004>
- Rudman, L., & Goodwin, S. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, 87, 494–509. <https://doi.org/10.1037/0022-3514.87.4.494>
- Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, 28, 61–79. <https://doi.org/10.1016/j.riob.2008.04.003>
- Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, 19(2), 109–111. <https://doi.org/10.1111/j.1467-9280.2008.02054.x>
- Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success.

- Sex Roles*, 61(9–10), 644–652. <https://doi.org/10.1007/s11199-009-9658-9>
- Schneider, B., Klager, C., Chen, I.-C., & Burns, J. (2016). Transitioning into adulthood: Striking a balance between support and independence. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 106–113. <https://doi.org/10.1177/2372732215624932>
- Schneider, M. C., Bos, A. L., & DiFilippo, M. (2021). Gender role violations and voter prejudice: The agentic penalty faced by women politicians. *Journal of Women, Politics & Policy*, 43(2), 117–133. <https://doi.org/10.1080/1554477X.2021.1981095>
- Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Processes*, 101(2), 230–242. <https://doi.org/10.1016/j.obhdp.2006.06.002>
- Shen, X., & Ferguson, M. J. (2021). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, 97, 104219. <https://doi.org/10.1016/j.jesp.2021.104219>
- Sherman, G. D., & Haidt, J. (2011). Cuteness and disgust: The humanizing and dehumanizing effects of emotion. *Emotion Review*, 3(3), 245–251. <https://doi.org/10.1177/1754073911402396>
- Shiramizu, V. K. M., Batres, C., Jones, A. L., Stern, J., DeBruine, L. M., & Jones, B. (2024). Who looks competent for what? Context-sensitivity in the predictors of competence judgments of face. <https://doi.org/10.31234/osf.io/eh52z>
- Smith, J. L., Paul, D., & Paul, R. (2007). No place for a woman: Evidence for gender bias in evaluations of presidential candidates. *Basic and Applied Social Psychology*, 29(3), 225–233. <https://doi.org/10.1080/01973530701503069>
- Spisak, B. R., Homan, A. C., Grabo, A., & Van Vugt, M. (2012). Facing the situation: Testing a biosocial contingency model of leadership in intergroup relations using masculine and feminine faces. *Leadership Quarterly*, 23(2), 273–280. <https://doi.org/10.1016/j.leaqua.2011.08.006>
- Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, 106(2), 186–208. <https://doi.org/10.1111/bjop.12085>
- Takinami, E. P. (2016). *Feminized Asians and masculinized Blacks: The construction of gendered races in the United States* [Senior thesis, University of Vermont]. <https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1126&context=hcoltheses>
- Walker, M., & Wänke, M. (2017). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PLOS ONE*, 12(10), e0181306. <https://doi.org/10.1371/journal.pone.0181306>
- Wen, F., Zuo, B., Ma, S., Xu, Y., Coley, J. D., & Wang, Y. (2020). Do we see masculine faces as competent and feminine faces as warm? Effects of sexual dimorphism on facial perception. *Evolutionary Psychology*, 18(4), 147470492098064. <https://doi.org/10.1177/1474704920980642>
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl, Vol. 1: Matters of public interest; Vol. 2: Personality and psychopathology* (pp. 89–113). University of Minnesota Press.
- 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>
- Wilson, J. P., Remedios, J. D., & Rule, N. O. (2017). Interactive effects of obvious and ambiguous social categories on perceptions of leadership: When double-minority status may be beneficial. *Personality and Social Psychology Bulletin*, 43(6), 888–900. <https://doi.org/10.1177/0146167217702373>
- Yee, A. H. (1973). Are the schools a “feminized society”? *Educational Leadership*, 31(2), 128–133.

SUPPLEMENTAL MATERIALS

Experiment 1

Results

Main Effects

A Trait effect showed higher Competence (Estimate = 4.53, 95% CI [4.38, 4.67]) relative to Dominance (Estimate = 4.39, 95% CI [4.25, 4.54]), $b = 0.133$, $SE = 0.02$, $z = 6.42$, $p < .001$, and Warmth (Estimate = 4.21, 95% CI [4.06, 4.35]), $b = 0.32$, $SE = 0.02$, $z = 15.47$, $p < .001$, ratings. Dominance versus Warmth ratings were also higher, $b = 0.187$, $SE = 0.02$, $z = 9.05$, $p < .001$. A Face Gender effect showed higher ratings of women's (Estimate = 4.52, 95% CI [4.35, 4.69]) relative to men's (Estimate = 4.23, 95% CI [4.06, 4.40]) faces. A Facial Characteristics effect showed higher ratings of feminized (Estimate = 4.40, 95% CI [4.26, 4.55]) relative to masculinized (Estimate = 4.35, 95% CI [4.20, 4.50]) faces. A Perceiver Gender effect showed higher ratings from women (Estimate = 4.49, 95% CI [4.29, 4.69]) relative to men (Estimate = 4.26, 95% CI [4.09, 4.43]).

Main Effects

A Face Gender effect showed that women's (Estimate = 4.43, 95% CI [4.22, 4.64]) versus men's (Estimate = 3.96, 95% CI [3.75, 4.17]) faces were endorsed higher in outcomes. A Facial Characteristics effect showed that feminized (Estimate = 4.26, 95% CI [4.09, 4.44]) versus masculinized (Estimate = 4.13, 95% CI [3.95, 4.30]) faces were endorsed higher in outcomes.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

For each evaluation, we regressed ratings on Facial Characteristics (−1 = masculinized, 1 = feminized), Face Gender (−1 = man, 1 = woman), Perceiver Gender (−1 = man, 1 = woman) and their interactions. The models included random effects structures such that intercepts varied

by participant and face identity. Facial Characteristics effects also varied by participant and face identity. See Table 4a-c in the main text for all model coefficient information.

Candidate endorsement

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3a in the main text). Participants were more likely to endorse women with feminized (Estimate = 4.49, 95% CI [4.26, 4.71]) relative to masculinized (Estimate = 4.25, 95% CI [4.02, 4.48]) characteristics, $b = 0.23$, $SE = 0.05$, $z = 4.30$, $p = .001$. No difference emerged for men with feminized (Estimate = 3.87, 95% CI [3.64, 4.10]) relative to masculinized (Estimate = 3.82, 95% CI [3.59, 4.05]), characteristics, $b = 0.05$, $SE = 0.05$, $z = 0.92$, $p = .80$.

Face Gender and Participant Gender interacted (Figure 4a). Women had higher endorsements of women candidates (Estimate = 4.45, 95% CI [4.17, 4.73]) relative to men candidates (Estimate = 4.00, 95% CI [3.72, 4.28]) candidates, $b = 0.45$, $SE = 0.13$, $z = 3.45$, $p = .003$. This difference was stronger among men (Estimate_{women} = 4.29, $SE = 0.13$; Estimate_{men} = 4.00, $SE = 0.14$), $b = 0.60$, $SE = 0.13$, $z = 4.73$, $p < .001$.

Likelihood of success

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3b). Participants evaluated women with feminized (Estimate = 4.61, 95% CI [4.39, 4.82]) relative to masculinized (Estimate = 4.41, 95% CI [4.20, 4.62]) characteristics as more likely to do well, $b = 0.20$, $SE = 0.05$, $z = 4.12$, $p = .002$. No difference emerged between endorsements of men candidates with feminized (Estimate = 4.05, 95% CI [3.83, 4.26]) relative to masculinized (Estimate = 3.98, 95% CI [3.78, 4.19]), characteristics, $b = 0.06$, $SE = 0.05$, $z = 1.30$, $p = .57$.

Face Gender and Participant Gender interacted (Figure 4b). Women participants evaluated women (Estimate = 4.59, 95% CI [4.34, 4.85]) relative to men (Estimate = 4.18, 95% CI [3.92, 4.43]) as more likely to do well, $b = 0.42$, $SE = 0.12$, $z = 3.35$, $p = .004$. This difference was stronger among men (Estimate_{women} = 4.42, $SE = 0.12$; Estimate_{men} = 3.85, $SE = 0.12$), $b = 0.57$, $SE = 0.12$, $z = 4.69$, $p < .001$.

Liking by other people

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3c). Participants thought other people would like reporting to women with feminized (Estimate = 4.56, 95% CI [4.34, 4.77]) relative to masculinized (Estimate = 4.28, 95% CI [4.07, 4.49]) characteristics, $b = 0.28$, $SE = 0.06$, $z = 4.90$, $p < .001$. No difference emerged between expected liking of men with feminized (Estimate = 4.01, 95% CI [3.79, 4.23]) relative to masculinized (Estimate = 4.02, 95% CI [3.81, 4.23]) characteristics, $b = -0.01$, $SE = 0.06$, $z = -0.19$, $p = .99$.

Face Gender and Participant Gender interacted (Figure 4c). Women thought others would like reporting to women (Estimate = 4.50, 95% CI [4.23, 4.76]) more than men (Estimate = 4.21, 95% CI [3.95, 4.47]), $b = 0.29$, $SE = 0.12$, $z = 2.40$, $p = .077$. This difference was stronger among men (Estimate_{women} = 4.34, $SE = 0.12$; Estimate_{men} = 3.82, $SE = 0.12$), $b = 0.52$, $SE = 0.12$, $z = 4.41$, $p = .001$.

Experiment 2

Results

Main Effects

Main effects of Trait and Face Gender replicated Experiment 1. The Trait effect showed faces rated higher in competence (Estimate = 4.64, 95% CI [4.52, 4.75]) versus dominance (Estimate = 4.47, 95% CI [4.35, 4.58]), $b = 0.17$, $SE = 0.01$, $z = 12.81$, $p < .001$, and warmth (Estimate = 4.26, 95% CI [4.15, 4.37]), $b = 0.37$, $SE = 0.01$, $z = 28.18$, $p < .001$. Dominance versus warmth ratings were also higher, $b = 0.20$, $SE = 0.01$, $z = 15.37$, $p < .001$. A Face Gender effect showed higher ratings of women's (Estimate = 4.58, 95% CI [4.44, 4.71]) versus men's (Estimate = 4.33, 95% CI [4.20, 4.47]) faces. Unique to Experiment 2, a Job Description effect showed higher ratings for candidates for the interpersonal (Estimate = 4.54, 95% CI [4.40, 4.68]) versus analytical (Estimate = 4.37, 95% CI [4.23, 4.50]) job.

Main Effects

Effects of Face Gender and Facial Characteristics replicated Experiment 1. A Face Gender effect showed higher endorsements of women (Estimate = 4.41, 95% CI [4.24, 4.58]) versus men (Estimate = 4.02, 95% CI [3.85, 4.19]). A Facial Characteristics effect showed higher endorsements of feminized (Estimate = 4.26, 95% CI [4.12, 4.40]) versus masculinized (Estimate = 4.17, 95% CI [4.03, 4.31]) faces. Unique to Experiment 2, a Job Description effect showed higher endorsements of candidates for the interpersonal (Estimate = 4.31, 95% CI [4.14, 4.48]) versus analytical (Estimate = 4.12, 95% CI [3.95, 4.28]) job.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

We regressed ratings on Job Description (-1 = analytical, 1 = interpersonal), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman),

Perceiver Gender ($-1 = \text{man}$, $1 = \text{woman}$) and their interactions. The models included random effects structures such that intercepts varied by participant and face identity. The random effects structure allowed facial characteristics effects to vary for each participant and for each face identity.

Candidate endorsement

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 6a). Participants were more likely to endorse higher endorse women with feminized (Estimate = 4.44, 95% CI [4.26, 4.63]) relative to masculinized (Estimate = 4.27, 95% CI [4.08, 4.45]), characteristics, $b = 0.18$, $SE = 0.05$, $z = 3.81$, $p = .001$. No difference emerged between endorsements of men with feminized (Estimate = 3.93, 95% CI [3.75, 4.11]) relative to masculinized (Estimate = 3.93, 95% CI [3.74, 4.12]), characteristics, $b = -0.003$, $SE = 0.05$, $z = -0.06$, $p = .99$.

Job Description qualified the interaction between Face Gender and Facial Characteristics (Figure 7). We characterized this interaction by examining Facial Characteristics effects for each Face Gender in each Job Description. For the analytical description, participants were more likely to endorse women candidates within feminized (Estimate = 4.33, 95% CI [4.12, 4.53]) relative to masculinized (Estimate = 4.13, 95% CI [3.91, 4.34]), characteristics, $b = 0.20$, $SE = 0.05$, $z = 3.71$, $p = .01$. No difference emerged between endorsements of men with feminized (Estimate = 3.79, 95% CI [3.58, 3.99]) relative to masculinized (Estimate = 3.85, 95% CI [3.63, 4.06]), characteristics, $b = -0.06$, $SE = 0.05$, $z = -1.17$, $p = .94$.

For the interpersonal description, participants were non-significantly more likely to endorse women candidates with feminized (Estimate = 4.56, 95% CI [4.35, 4.78]) relative to masculinized (Estimate = 4.40, 95% CI [4.18, 4.63]), characteristics, $b = 0.16$, $SE = 0.06$, $z =$

2.84, $p = .09$. No difference emerged between endorsements of men with feminized (Estimate = 4.07, 95% CI [3.86, 4.28]) relative to masculinized (Estimate = 4.01, 95% CI [3.79, 4.24]), faces, $b = 0.06$, $SE = 0.06$, $z = 1.03$, $p = .97$.

Likelihood of success

Face Gender and Facial Characteristics effects were qualified by their interaction (Figure 6b). Participants evaluated women with feminized (Estimate = 4.57, 95% CI [4.39, 4.74]), relative to masculinized (Estimate = 4.41, 95% CI [4.24, 4.58]), characteristics as more likely to do well, $b = 0.16$, $SE = 0.04$, $z = 3.95$, $p < .001$. No difference emerged between endorsements of men with feminized (Estimate = 4.08, 95% CI [3.91, 4.26]) relative to masculinized (Estimate = 4.09, 95% CI [3.91, 4.26]), characteristics, $b = -0.01$, $SE = 0.04$, $z = -0.15$, $p = .99$.

Liking by other people

Face Gender and Facial Characteristics effects were qualified by their interaction (Figure 6c). Participants thought that other people would like reporting to women with feminized (Estimate = 4.50, 95% CI [4.33, 4.67]), relative to masculinized (Estimate = 4.29, 95% CI [4.11, 4.47]), characteristics, $b = 0.21$, $SE = 0.04$, $z = 4.77$, $p < .001$. No difference emerged between expected liking of men with feminized (Estimate = 4.04, 95% CI [3.87, 4.22]) relative to masculinized (Estimate = 4.04, 95% CI [3.86, 4.22]), characteristics, $b = 0.002$, $SE = 0.04$, $z = 0.04$, $p > .99$.

Face Gender and Participant Gender interacted (Figure 8). Women thought others would like reporting to women (Estimate = 4.35, 95% CI [4.15, 4.55]) more than men (Estimate = 4.04, 95% CI [3.83, 4.24]), $b = 0.31$, $SE = 0.10$, $z = 3.03$, $p = .01$. This difference was stronger among men (Estimate_{women} = 4.44, 95% CI [4.25, 4.63]; Estimate_{men} = 4.05, 95% CI [3.86, 4.24]), $b = 0.39$, $SE = 0.10$, $z = 3.78$, $p < .001$.

Job Description effects on College Student Beliefs

We tested whether being exposed to the different job descriptions in the task affected gender-biased beliefs about college men and women. No differences in gender biased beliefs about college students' warmth, competence, dominance, and responsibility emerged by job description, $ps > .09$.

Face Gender by Facial Characteristics Contrasts

Experiment 1

<i>Face Gender to Facial Characteristics contrast</i>	<i>estimate</i>	<i>CI</i>	<i>p</i>
Female Feminized - Male Feminized	0.57	0.32, 0.83	< .001
Female Feminized - Female Masculinized	0.24	0.14, 0.33	< .001
Female Feminized - Male Masculinized	0.61	0.36, 0.85	< .001
Male Feminized - Female Masculinized	-0.34	-0.58, -0.09	.070
Male Feminized - Male Masculinized	0.03	-0.06, 0.13	.984
Female Masculinized - Male Masculinized	0.37	0.14, 0.60	.021

Experiment 2

<i>Face Gender to Facial Characteristics contrast</i>	<i>estimate</i>	<i>CI</i>	<i>p</i>
Female Feminized - Male Feminized	0.49	0.28, 0.70	< .001
Female Feminized - Female Masculinized	0.18	0.10, 0.26	< .001
Female Feminized - Male Masculinized	0.48	0.27, 0.70	< .001
Male Feminized - Female Masculinized	-0.30	-0.52, -0.09	.054
Male Feminized - Male Masculinized	-0.00	-0.08, 0.08	1.000
Female Masculinized - Male Masculinized	0.30	0.09, 0.51	.054