



Trait inferences from the “big two” produce gendered expectations of facial features[☆]

Hayley A. Liebenow^{a,*}, Kathryn L. Boucher^b, Brittany S. Cassidy^a

^a University of North Carolina at Greensboro, USA

^b University of Indianapolis, USA

ARTICLE INFO

Keywords:

Gender stereotypes
Face perception
Impression formation
Sexual dimorphism

ABSTRACT

Prescriptive stereotypes based on, respectively, agency and communality reflect how people expect men and women to behave. Deviating from such prescriptions limits opportunities for men and women in ways that reinforce traditional gender roles. In the current work, we examine whether people have expectations of gendered facial features based on agentic and communal descriptions of targets and if these expectations extend to who people think is best suited for workplace tasks. Across five experiments, people expected more facial masculinity for targets paired with agentic relative to communal traits (Experiments 1, 2a-b) and workplace behaviors (Experiments 3a-b). This expectation effect emerged when gendered facial features (e.g., more masculinized and feminized versions of face identities) were manipulated across (Experiment 1) and within (Experiments 2a-b, 3a-b) gender, regardless of whether traits were explicitly stated (Experiments 1, 2a-b, 3a) or inferred (Experiment 3b), and regardless of trait valence. When people made decisions about two same-gender faces, the gender of those faces accentuated trait effects. More masculine male (relative to female) faces were consistently expected *more* for agentic traits and workplace tasks, but consistently expected *less* for communal traits and workplace tasks (Experiments 2a, 3a-b). We then conceptually replicated expectation effects by showing that mental representations of agentic and communal faces appear correspondingly gendered (Experiment 4). Finally, we provide exploratory analyses showing that expectation effects may differentially vary by perceiver gender across contexts. These findings illustrate a non-verbal route by which people make decisions based on gender stereotypes that have wide-ranging implications for workplace behavior.

Americans have expectations about who men and women are and who they should be (Eagly & Crowley, 1986; Prentice & Carranza, 2002). Decades of research on descriptive and prescriptive gender stereotypes (e.g., Eagly, Nater, Miller, Kaufmann & Sczesny, 2020) have focused on observations that such stereotypes are tied to traditional beliefs about social roles (Diekmann & Eagly, 2000; Prentice & Carranza, 2002). People believe that women should possess communal traits reflecting a relationship orientation (e.g., warm) and that men should possess agentic traits characteristic of leadership (e.g., dominant). Deviating from these expectations can be devastating. For example, people deviating from gender expectations receive lower popularity ratings from peers (Costrich, Feinstein, Kidder, Marecek & Pascale, 1975). Moreover, women who do versus do not deviate from gender prescriptions receive unfair treatment in sex discrimination incidents (Burgess & Borgida, 1999). Gaining a better understanding of how

expectations of men and women are reflected in social cognition is thus essential to understand bias against people who are not what they are expected to be.

To date, the makeup and consequences of gender stereotypes have largely been examined through self-reported beliefs (e.g., Eagly et al., 2020) and evaluations of resumes or vignettes (e.g., Phelan, Moss-Racusin & Rudman, 2008; Rudman & Glick, 1999). Inferences reflecting traditional social roles, however, also emerge from non-verbal cues like faces (e.g., Wen et al., 2020). Gendered expectations of facial appearance may thus underlie differences in expectations for behaviors and tasks for which people are believed to be best suited. We explore this possibility by examining how gendered inferences from communion and agency affect expectations for how men and women should look. Specifically, we examine this possibility through the lens of face impressions in various contexts with a variety of paradigms.

[☆] This paper has been recommended for acceptance by Dr Rachael Jack.

* Corresponding author at: Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart, PO Box 26180, Greensboro, NC 27412, USA.

E-mail address: haliebenow@uncg.edu (H.A. Liebenow).

Faces elicit impressions related to agency and communality (i.e., the “big two”) through sexually dimorphic features (Cassidy & Liebenow, 2021; Walker & Wänke, 2017; Wen et al., 2020) that affect how people evaluate (Sutherland, Young, Mootz & Oldmeadow, 2015) and behave (Hehman, Flake & Freeman, 2015) toward targets. For example, women with more masculinized relative to feminized features are evaluated as being more agentic and less communal (e.g., Cassidy & Liebenow, 2021) and as having less credible sexual harassment allegations (Goh, Bandt-Law, Cheek, Sinclair & Kaiser, 2021). Although people largely agree on face impressions, such impressions do not strongly relate to actual behaviors (e.g., Rule, Krendl, Ivcevic & Ambady, 2013). Consensual, but potentially incorrect, gendered trait expectations of faces may thus limit the opportunities afforded to people. For example, expectations for how employees should look based on assigned tasks may result in more job offers toward people with expectation-matching faces. Here, we define sexually dimorphic facial features as the extent to which facial features appear masculinized or feminized (i.e., gendered). Thus, the term “gendered” refers to facial features reflecting such sexually dimorphic facial features. We manipulated these features in two ways. First, to the extent that gender category exemplars have sexually dimorphic features more representative of that category, we morph these faces across the gender continuum (i.e., man to woman; Experiment 1). Second, we manipulate these features using a widely used morphing software that manipulates faces specifically across this dimension (Experiments 2a–3b).

We propose that people will expect others with agentic and communal traits to have, respectively, more masculinized and feminized facial features. We also anticipate expectations to be reflected in the faces selected as best suited for agentic and communal workplace tasks. Establishing a link between gendered traits and expectations of facial appearance will forge a link between currently disparate literatures on gender stereotypes (e.g., Prentice & Carranza, 2002) and face perception (e.g., Sutherland et al., 2015) in that expectations of facial features may be non-verbal route to gender role reinforcement (e.g., Martin & Slepian, 2020). To this end, we first summarize work on gender stereotypes and agentic and communal trait inferences. We then discuss work on face impressions, focusing on relations between masculinized and feminized facial features and impressions of these traits. Finally, we present six experiments using a variety of paradigms to show that people have gendered expectations of facial features based on traits and workplace contexts.

1. Associations between gender stereotypes and the “big two”

Gender stereotypes can be descriptive by reflecting qualities people ascribe to men and women (for a review, see Koenig, 2018) and prescriptive by reflecting beliefs of what men and women should be (e.g., Prentice & Carranza, 2002). These stereotypes have consequences that enforce traditional social roles (for a review, see Eagly & Wood, 2012). Beliefs about what behaviors are appropriate for men and women, for example, relate to differences in how men and women display aggressive and prosocial behavior (Eagly & Wood, 1991). Initial research identifying the content of these stereotypes yielded traits characterizing what people think men and women are and should be (Bem, 1974). Work conducted almost 30 years later showed the persistence of these stereotype-characterizing traits (Auster & Ohm, 2000).

A core pattern in this work is that emergent traits largely parallel the “big two” traits of social cognition (for reviews, see Bruckmüller & Abele, 2013; Martin & Slepian, 2020; Wiggins, 1991). Whereas people believe that men are agentic and should possess traits characterizing their agency, they believe that women are communal and should possess traits characterizing their communality. Although some work suggests dynamic gender stereotypes (Diekmann & Eagly, 2000; Eagly et al., 2020), work summarizing 73 years of public opinion polls showed pervasive gender stereotypes reflecting the big two (Eagly et al., 2020). Even though people may assume changes in gender stereotypes based on

societal shifts (e.g., more women obtaining political offices, Bonneau & Kanthak, 2020), gender prescriptions remain pervasive in the United States. Reflecting their pervasive nature, theoretical work has asserted that gender prescriptions are key to how people experience the world (Martin & Slepian, 2020). For example, people’s earliest social-cognitive processes seem to be organized by gender (e.g., categorization; Bem, 1981; Mackie, Hamilton, Susskind, & Rosselli, 1996).

Because gender is linked to the big two (e.g., Martin & Slepian, 2020), it is perhaps unsurprising that deviating from gender prescriptions elicits negative consequences. Consequences for women have been the subject of much research. For example, role congruity theory asserts that women are unfavorably evaluated for leadership in part because evaluations of agentic behaviors fulfilling the prescriptions of leadership roles are less favorable when enacted by women than men (Eagly & Karau, 2002). Indeed, women described with agentic relative to communal traits are deemed less hireable for managerial roles (Phelan et al., 2008). Men also encounter negative consequences for deviating from gender prescriptions. For example, men are evaluated as less qualified than women for female-stereotyped jobs (Davison & Burke, 2000).

Research on gender stereotypes and their relation to the big two has largely been conducted using verbal measures (e.g., Eagly et al., 2020). For example, people might be asked to rate how desirable it is for a man or a woman to possess different trait characteristics on a scale (e.g., Prentice & Carranza, 2002). Notably, people also have trait impressions reflecting the “big two” based on gendered facial features (e.g., masculinized and feminized featural content, Walker & Wänke, 2017). Paralleling work using verbal measures, people may expect others to look a certain way based on their traits, thus affecting the opportunities afforded to them via a non-verbal route.

2. Facial features elicit impressions of the “big two”

Trait impressions from faces are enduring (e.g., Ambady & Rosenthal, 1992) and often override other incoming information (e.g., Jaeger, Evans, Stel & van Beest, 2019). Although face impressions are largely consensual, they do not often accurately predict targets’ behavior (Rule et al., 2013). Although many facial features have been examined regarding their connection to such consensual impressions (for a review, see Zebrowitz & Montepare, 2008), we focus on masculinized and feminized facial features given their direct connection to gender and its known consequences for how people evaluate faces. For example, people more positively evaluate more versus less gender-typical faces (Oh, Buck, & Todorov, 2019; Sutherland et al., 2015).

Masculinized and feminized facial features seem to track impressions of the big two. For example, people evaluated a masculinized relative to a feminized version of Kamala Harris’s face as reflecting more agentic and fewer communal traits (Cassidy & Liebenow, 2021). That these impressions came from the same identity manipulated on masculinity suggests that sexually dimorphic features rather than other identity-specific features affected impressions. Likewise, models have shown higher facial masculinity to be associated with agency-related competence stereotypes causing faces manipulated to appear more competent to be more likely to be categorized as male (Oh et al., 2019). The associations between masculinized and feminized facial features and the big two has been shown using a variety of methods. Several experiments have shown it via widely-used explicit trait ratings tasks (e.g., Cassidy & Liebenow, 2021; Walker & Wänke, 2017). These associations have also been found at an implicit level (Wen et al., 2020), which suggests their automaticity.

Consequences for faces varying sexually dimorphic features seem to parallel the negative consequences shown for people who deviate from gender prescriptions. For example, people representing Kamala Harris’s face as more masculine were less likely to support her in 2020 (Cassidy & Liebenow, 2021), a finding complementing work showing that actual female politicians with more masculinized faces receive fewer votes

(Hehman, Carpinella, Johnson, Leitner & Freeman, 2014). Men also experience negative consequences based on their features. For example, men with *more* masculine faces are described as colder than men with more feminine faces (Walker & Wänke, 2017). Because coldness is negatively related to approachability (Perrine, 1998), it suggests that men with more masculine relative to more feminine faces will be deemed less approachable, which likely limits what social interactions they are afforded.

Although this work has established that sexually dimorphic facial features relate to impressions of the big two, there are gaps in the literature still to address. First, associations have been established using few faces (Cassidy & Liebenow, 2021; Walker & Wänke, 2017; Wen et al., 2020). Establishing associations across myriad faces is important to resolve that past findings reflect generalizable associations rather than ones specific to a few identities. Second, although this work has shown *reactions* to faces, it has not directly addressed *expectations* of faces. Some work indirectly supports expectations of sexually dimorphic features based on gendered traits. For example, people have more feminized image representations of language arts than physics teachers' faces (Degner, Mangels & Zander, 2019), which is consistent with people's perceptions of STEM fields as non-communal (Brown, Thoman, Smith, & Diekmann, 2015). Moreover, some work (e.g., Degner et al., 2019; Imhoff, Woelki, Hanke & Dotsch, 2013) has assessed how people mentally represent the "big two" in faces (e.g., Oliveira, Garcia-Marques, Dotsch & Garcia-Marques, 2019). Building on these findings, the present work provides a connection between gendered traits to expectations of gendered facial features by showing a direct connection across a variety of contexts. Showing this direct connection is important because it would directly link work on face impressions to work on prescriptive gender stereotypes (Prentice & Carranza, 2002). Here, we examined expectations of sexually dimorphic facial features based on agentic or communal trait or workplace descriptions across a variety of face identities.

3. Who has expectations of facial appearance? Exploring potential perceiver gender effects

We anticipated that people would have expectations of more masculinized and more feminized facial features based on, respectively, described agency and communality. An open question regarded if some people would have these expectations more than others. This question is important to consider because it suggests that the extent to which people are afforded opportunities based on sexually dimorphic facial features may depend on characteristics of the decision maker. Although many factors likely affect these expectations, we focused exploratory analyses on perceiver gender.

People who hold societal privilege are largely motivated to uphold it (Eagly & Karau, 2002; Eagly & Wood, 1991; Pleasants, 2011). Because men hold more societal privilege than women in terms of having leadership characterized by agency (e.g., Bosak & Sczesny, 2011; Koenig, Eagly, Mitchell & Ristikari, 2011), one theory is that that men maintain this status quo by reinforcing gender roles (Swami et al., 2013; Vandello & Bosson, 2013). Indeed, traditional gender roles benefit men in terms of preserving their societal power (Glick & Fiske, 1996). Supporting this idea, greater gender role reinforcement by men relative to women has been shown both experimentally (Ho et al., 2015; Rudman, Moss-Racusin, Phelan & Nauts, 2012) and through tendencies by men to score higher than women on measures linked to traditional social role endorsement (Eagly & Karau, 2002; Fiske, Cuddy, Glick & Xu, 2002; Glick et al., 2000; Rudman & Glick, 2001). We reasoned that these patterns may translate to men having stronger expectations of masculinized and feminized facial features than women based on the big two.

4. Current research

We tested whether people expect faces described by agentic and

communal traits and tasks to appear, respectively, more masculinized and more feminized. We established these expectations by having people select a face from an array altered from one male to one female exemplar reflecting their expectations of a person described by agentic or communal traits (Experiment 1). Second, we replicated and extended this finding by testing whether people have such expectations both between and within face gender (Experiments 2a-b). These experiments also used positive (Experiment 2a) and negative (Experiment 2b) traits reflecting the big two, which established that expectations of facial masculinity and femininity generalize across valenced traits. Third, we tested whether gendered expectations of facial masculinity and femininity generalize from trait to workplace descriptions (Experiments 3a-b). These experiments tested whether expectations of facial masculinity and femininity align with gender prescriptions through the accolades people receive (Experiment 3a) and their assigned tasks (Experiment 3b). Finally, we used reverse correlation to conceptually replicate the prior experiments using a more implicit and open-ended approach (Experiment 4). Across experiments and contexts, we provide consistent evidence that people have expectations of sexually dimorphic facial features based on the big two. Although expectations of facial features based on gendered concepts has been studied in related research (i.e., job occupation; Degner et al., 2019), the current work builds a novel and direct case that people not only have these expectations of what an "agentic" or "communal" person looks like (Experiment 1 and 4), but that these expectations reflect sexually dimorphic facial features (Experiments 2a – 4) and apply when people consider traits (Experiments 2a-b) and common workplace situations (Experiments 3a-b). We also provide exploratory evidence that these expectations of facial masculinity and femininity based on the big two seems pronounced among men perceivers, but that men and women perceivers do not differentially show this effect when evaluating workplace contexts.

5. Experiment 1

Experiment 1 established expectations of sexually dimorphic facial features reflecting the big two. People saw face arrays altered on a continuum from a male to a female face. They selected the face reflecting a target described by agentic or communal traits. People react to relative facial masculinity and femininity by forming impressions reflecting the big two (e.g., Wen et al., 2020). If such patterns extend to *expectations* of sexually dimorphic facial features, we hypothesized that people would expect a target described by agentic relative to communal traits to appear more masculine. Given that men often enforce traditional gender roles more than women (Eagly & Karau, 2002; Glick & Fiske, 1996; Rudman & Glick, 1999, 2001), we explored if this pattern was pronounced for men relative to women.

5.1. Method

5.1.1. Participants

Two hundred and twenty-five participants from MTurk participated ($M_{age} = 37.96$ years, $SD = 12.12$; $M_{years\ of\ education} = 15.34$, $SD = 2.52$; 101 identifying as female). We targeted a sample of 200 given work showing listed traits to affect the extent to which traits are reflected in faces (Cassidy & Krendl, 2018). We oversampled to account for anticipated exclusions (e.g., failing attention checks), although all passed. Of the 225 participants, 188 identified as White, 11 as Black, 18 as Asian, 7 as multi-racial, and 1 as American Indian/Alaska Native. Of the 225 participants, 213 identified as non-Hispanic. Across experiments, all participants provided informed consent and were compensated \$1.00. All experiments were approved by the university's IRB. Across experiments, all measures, manipulations, and exclusions have been disclosed.

5.1.2. Stimuli

To measure expectations of sexually dimorphic facial features, we created stimuli following procedures from past work (Cassidy & Krendl,

2018; Cole, Trope & Balcetis, 2016; Epley & Whitchurch, 2008) and that are summarized here. Two male and two female neutrally expressive White faces were selected from the Chicago Face Database (CFD; Ma, Correll & Wittenbrink, 2015). We created two male-female face pairs of comparable age and attractiveness from these faces (named in the database as WM-204 and WF-208; WM-029 and WF-242) similar in attractiveness and age based on database norms. See the Supplemental Materials for CFD masculinity and femininity norms, among others.

We cropped face images below the chin and at the top of the forehead to remove hair and features beyond the face itself. We grayscaled each face and applied a 30% blur to each so unaltered faces would not appear clearer than the altered faces. We used Abrasoft Fantamorph to generate 11 altered face images from each pair of original face images. The alteration procedure involves aligning two faces by matching points on the one face with identical points on the other. After alignment, we altered faces by changing the percentage of the one face represented in the other face to create a continuum. We altered the faces of each pair in 10% increments, yielding 11 faces ranging from 0% male (i.e., 100% female face and 0% male face) to 100% male (i.e., 0% female face and 100% male face). The altered face image at the continuum center thus comprised 50% each of female and male faces. See the Supplemental Material for more information on viewing conditions.

5.1.3. Procedure

Each participant completed a task comprised of two blocks. They read, “You will read some information about two people: Person 1 and Person 2. Please read the information carefully. You will then be asked to choose which picture from an array best resembles that person based on what you know about them. After making six decisions about each person, we will ask you to complete three short questionnaires and to provide some demographic information.” Prior work has used a similar number of decisions to assess facial appearance expectations (e.g., Cassidy & Krendl, 2018; Cole et al., 2016).

The blocks differed by whether the target was described by agentic (assertive, ambitious, independent, strong, determined, and persistent) or communal (cooperative, warm, friendly, caring, compassionate, and sympathetic) traits. Each target was only described by one trait type, and

participants saw targets described by all agentic and all communal traits. Traits were selected from a database of gendered trait inferences (Diehl, Owen & Youngblade, 2004) and are widely used examples of the big two (e.g., Abele, 2003; Bruckmüller & Abele, 2013; Gervais & Hillard, 2011; Trapnell & Paulhus, 2012). Each block comprised six randomly presented trials per target. Each trial used a different trait. For each trial, people read, “After taking an in-depth personality test, Person 1 [2] was evaluated with a variety of personality traits. Below, you will see one of the traits that Person 1 [2] was rated very highly on. The test showed that Person 1 [2] is: [Trait]. Knowing that Person 1 [2] is described as [trait], which one of the photos below do you think is Person 1 [2]?” Below this question was the above-described continuum of faces displayed in a random order (Fig. 1b). Participants selected which face best represented the target six times within each block.

At the end of each block, people rated the target’s gender stereotypicality using a scale from -3 (*very feminine*) to 3 (*very masculine*). Four versions counterbalanced whether agentic or communal traits were presented with each face pair and whether people selected faces described by agentic or communal traits first or second. After the task, people indicated whether the task involved people, food, pets, or furniture. All indicated people.

5.1.3.1. Participant characterization. Because we conducted exploratory analyses including perceiver gender, we characterized whether any gender differences emerged across attitudes and beliefs related to gender. Here, we characterized participants on political ideology, sexism, and social role endorsement in a random order after the task.

5.1.3.1.1. Political ideology. People indicated political ideology over four items (overall, economic issues, social issues, and foreign policy issues) on a scale from 1 (*extremely conservative*) to 9 (*extremely liberal*). Responses (Cronbach’s $\alpha = 0.96$) were averaged to create composite political ideology scores ($M = 5.83$, $SD = 2.37$).

5.1.3.1.2. Sexism. People 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 from 0 (*disagree*



Fig. 1. Female-male face pair morphed in 10% increments (A) and an example trial in Experiment 1 that including the face array displayed in a random order (B).

strongly) to 5 (agree strongly). Higher scores indicate more sexism. Items measuring hostile (Cronbach's $\alpha = 0.93$; $M = 1.46$, $SD = 1.14$) and benevolent (Cronbach's $\alpha = 0.92$; $M = 1.90$, $SD = 1.21$) sexism were averaged to create hostile and benevolent sexism scores.

5.1.3.1.3. *Social role beliefs.* People completed the Social Roles

Questionnaire (SRQ; Baber & Tucker, 2006). Participants indicated how much they agreed with each item (e.g., “For many important jobs, it is better to choose men instead of women”) using a scale from 0% to 100% in 10% increments. Higher scores reflect more traditional social role beliefs. Items were averaged to create social role belief scores

Table 1

Descriptive (M [SD]) and inferential statistics for participant characterization measures in Experiments 1, 2a, 2b, 3a, 3b, and 4.

a. Experiment 1	Men ($N = 124$)	Women ($N = 101$)	t	p	d [95% CI]
Hostile sexism	1.66 (1.11)	1.21 (1.14)	2.99	0.003	0.40 [0.13, 0.66]
Benevolent sexism	2.12 (1.16)	1.63 (1.21)	3.07	0.002	0.41 [0.15, 0.68]
Traditional gender role beliefs	32.26 (19.35)	25.15 (21.19)	2.63	0.009	0.35 [0.09, 0.62]
Political ideology	5.72 (2.16)	5.97 (2.41)	0.79	0.42	0.11 [−0.16, 0.37]
Age	36.99 (10.77)	39.16 (11.62)	1.45	0.15	0.19 [−0.07, 0.46]
Years of Education	15.52 (2.47)	15.13 (2.15)	1.24	0.22	0.17 [−0.10, 0.43]
b. Experiment 2a	Men ($N = 112$)	Women ($N = 85$)	t	p	d [95% CI]
Hostile sexism	1.91 (1.27)	1.32 (1.24)	3.39	< 0.001	0.47 [0.20, 0.75]
Benevolent sexism	2.07 (0.98)	1.62 (1.18)	3.03	0.003	0.42 [0.15, 0.70]
Traditional gender role beliefs	31.88 (16.91)	22.79 (18.98)	3.66	< 0.001	0.51 [0.23, 0.79]
Political ideology	4.61 (1.77)	4.72 (1.86)	−0.45	0.66	0.06 [−0.21, 0.34]
Age	37.84 (10.93)	45.67 (13.24)	−4.70	< 0.001	0.66 [0.38, 0.94]
Years of Education	15.57 (3.53)	15.84 (2.19)	−0.62	0.54	0.08 [−0.19, 0.36]
c. Experiment 2b	Men ($N = 112$)	Women ($N = 87$)	t	p	d [95% CI]
Hostile sexism	1.59 (1.35)	1.29 (1.16)	1.64	0.10	0.23 [−0.06, 0.66]
Benevolent sexism	1.86 (1.11)	1.70 (1.14)	0.99	0.32	0.14 [−0.16, 0.47]
Traditional gender role beliefs	28.03 (19.60)	25.39 (19.12)	0.95	0.34	0.14 [−2.83, 8.10]
Political ideology	4.73 (1.64)	4.64 (1.82)	0.39	0.69	0.06 [−0.39, 0.58]
Age	39.09 (11.57)	41.21 (12.56)	−1.23	0.22	0.18 [−5.50, 1.27]
Years of Education	15.23 (2.16)	15.33 (1.96)	−0.34	0.73	0.05 [−0.69, 0.48]
d. Experiment 3a	Men ($N = 113$)	Women ($N = 101$)	t	p	d [95% CI]
Hostile sexism	1.77 (1.25)	1.19 (1.02)	3.57	<0.001	0.49 [0.26, 0.90]
Benevolent sexism	1.97 (1.12)	1.83 (1.06)	1.70	0.09	0.23 [0.04, 0.56]
Traditional gender role beliefs	30.87 (19.14)	25.73 (18.98)	1.97	0.05	0.27 [−0.01, 10.28]
Political ideology	4.74 (1.59)	4.87 (1.73)	−0.57	0.57	0.08 [−0.58, 0.32]
Age	39.45 (11.72)	41.23 (11.89)	−1.10	0.27	0.15 [−4.96, 1.41]
Years of Education	15.59 (2.21)	15.43 (2.51)	0.52	0.60	0.07 [−0.47, 0.80]
Social dominance orientation	2.35 (1.44)	1.94 (1.30)	2.15	0.03	0.30 [0.03, 0.78]
Identity threat	1.89 (1.14)	2.95 (1.40)	−6.08	<0.001	0.83 [1.89, 2.95]
e. Experiment 3b	Men ($N = 95$)	Women ($N = 118$)	t	p	d [95% CI]
Hostile sexism	1.76 (1.25)	1.18 (1.02)	3.72	<0.001	0.51 [0.27, 0.88]
Benevolent sexism	1.97 (1.12)	1.83 (1.06)	0.94	0.35	0.13 [−0.15, 0.43]
Traditional gender role beliefs	28.15 (20.06)	26.31 (18.35)	0.70	0.49	0.10 [−3.36, 7.03]
Political ideology	4.63 (1.84)	4.59 (1.73)	0.16	0.87	0.02 [−0.44, 0.52]
Age	36.48 (9.86)	40.50 (11.46)	−2.70	0.01	0.37 [−6.95, 1.09]
Years of Education	15.18 (2.43)	15.36 (2.52)	−0.53	0.60	0.07 [−0.86, 0.49]
Social dominance orientation	2.28 (1.41)	2.08 (1.24)	1.11	0.27	0.15 [−0.16, 0.56]
Identity threat	1.84 (1.09)	3.06 (1.37)	−7.05	<0.001	−0.97 [−1.56, −0.88]
a. Experiment 4	Men ($N = 75$)	Women ($N = 44$)	t	p	d [95% CI]
Hostile sexism	1.88 (1.21)	1.85 (1.14)	0.13	0.90	0.02 [−0.42, 0.48]
Benevolent sexism	2.26 (1.04)	2.56 (1.01)	−1.52	0.13	0.29 [−0.69, 0.09]
Traditional gender role beliefs	39.24 (20.39)	39.53 (20.39)	−0.07	0.94	0.01 [−7.96, 7.38]
Political ideology	4.69 (1.51)	4.63 (1.79)	0.21	0.83	0.04 [−0.54, 0.67]
Age	39.41 (11.13)	38.25 (10.26)	0.57	0.57	0.11 [−2.90, 5.23]
Years of Education	15.01 (2.27)	15.70 (2.36)	−1.59	0.12	0.30 [−1.56, 0.17]
Social dominance orientation	2.72 (1.49)	2.80 (1.34)	−0.29	0.77	0.06 [−0.62, 0.46]

(Cronbach's $\alpha = 0.90$; $M = 29.07$, $SD = 20.46$).

People lastly provided demographic information.

5.2. Results

5.2.1. Gender differences emerged on participant characterization measures

Relative to women, men had more hostile sexism, benevolent sexism, and traditional social role beliefs. Men and women were similar in their political ideology, age, and years of education. See Table 1a for descriptive and inferential statistics.

5.2.2. Gendered traits elicited differential perceptions of gender stereotypicality

Agentic and communal traits elicited differential perceptions of gender stereotypicality. A one sample t -test against the scale midpoint (0) showed more masculine gender stereotypicality ($M = 0.34$, $SD = 1.43$) for targets described by agentic traits, $t(224) = 3.60$, $p < .001$, $d = 0.24$, 95% CI [0.11, 0.37]. A one sample t -test against the scale midpoint (0) showed more feminine gender stereotypicality ($M = -0.90$, $SD = 1.36$) for targets described by communal traits, $t(224) = 9.93$, $p < .001$, $d = 0.66$, 95% CI [0.52, 0.81]. A paired samples t -test showed that people perceived the target described by agentic relative to the communal traits as more stereotypically masculine, $t(224) = 8.92$, $p < .001$, $d = 0.59$, 95% CI [0.45, 0.74].

5.2.3. The big two elicited expectations of gendered facial features

5.2.3.1. Analytic plan. Across experiments, we fitted mixed-effects models using *lme4* (Bates, Mächler, Bolker & Walker, 2015) in R, estimating confidence intervals using the confint function. P -values were calculated using *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017). Estimated marginal means were obtained using *emmeans* (Lenth, Singmann, Love, Buerkner, & Herve, 2018). Here, we coded responses on each trial using the percentage of the male exemplar reflected in selected face (i.e., selection masculinity). The score for each trial could thus range from 0% (i.e., the 100% female face) to 100% (i.e., the 100% male face). Each participant had 12 scores (six per face pair). Selection masculinity was regressed on Trait Type ($-1 =$ agentic and $1 =$ communal), Perceiver Gender ($-1 =$ man, $1 =$ woman), and their interaction as fixed effects.

This model included a random effects structure such that intercepts were allowed to vary by participant and by the specific trait on each trial. We allowed a Trait Type effect to vary by participant. Accounting for variability from differences between the two face pairs by nesting random effects by trait within face pair did not change the below described results. We report unstandardized regression coefficients (B). We ran this model after standardizing the dependent variable to also report standardized regression coefficients (b). 95% CIs refer to analyses outputting unstandardized regression coefficients.

5.2.3.2. Expectations of gendered facial features. Across analyses in all experiments, "Estimate" refers to the estimated marginal mean obtained from *emmeans* output in R. As expected, a Trait Type effect showed that people selected more masculine faces to reflect targets described by agentic (Estimate = 49.80, 95% CI [44.60, 55.00]) relative to communal (Estimate = 41.10, 95% CI [36.00, 46.20]) traits, $B = -4.35$, $SE = 1.86$, $b = -0.12$, $t = 2.34$, $p = .03$, 95% CI [-7.96, -0.74]. A Participant Gender effect showed that men (Estimate = 48.20, 95% CI [44.40, 51.90]) selected more masculine faces than women (Estimate = 42.80, 95% CI [38.90, 46.70]), $B = -2.68$, $SE = 0.91$, $b = -0.08$, $t = 2.96$, $p = .003$, 95% CI [-4.46, -0.90].

The expected interaction qualified these effects, $B = 2.49$, $SE = 1.27$, $b = 0.07$, $t = 1.97$, $p = .05$, 95% CI [0.01, 4.98] (Fig. 2). Men selected more masculine faces to reflect targets described by agentic (Estimate = 55.00, 95% CI [49.20, 60.80]) relative to communal (Estimate = 41.30,

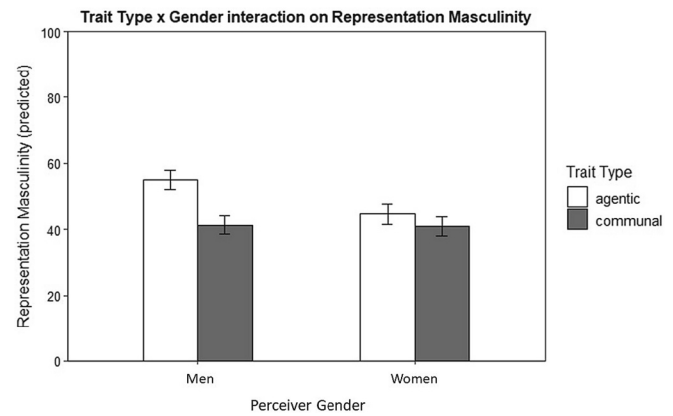


Fig. 2. An interaction between Trait Type and Perceiver Gender in Experiment 1 showed that men, but not women, were more likely to choose a more masculine morph as reflecting agentic relative to communal traits.

95% CI [35.60, 47.00]) traits, Estimated difference = 13.68, $SE = 4.35$, $t = 3.15$, $p = .003$. Women did not show this difference (Estimate_{agentic} = 44.70, 95% CI [38.50, 50.80]; Estimate_{communal} = 40.90, 95% CI [35.00, 46.90]), Estimated difference = 3.71, $SE = 4.64$, $t = 0.80$, $p = .43$. A sensitivity analysis indicated a minimum detectable interaction effect size of $b = 0.0905$ with power = 0.80 and alpha = 0.05. The effect size from the sensitivity analysis is the standardized regression coefficient.

5.3. Discussion

People expected targets described by agentic relative to communal traits to appear more masculine. This finding aligns with work showing that people associate agency and communality, with, respectively, masculinity and femininity (Bruckmüller & Abele, 2013; Diehl et al., 2004; Martin & Slepian, 2020). It also aligns with work on reactions to sexual dimorphism showing agentic and communal face impressions based on, respectively, gendered facial features (Walker & Wänke, 2017; Wen et al., 2020). This finding extends and connects these bodies of work. By showing expectations of gendered facial features based on agentic and communal descriptions, the current experiment links work on prescriptive gender stereotypes and facial features.

Exploratory analyses showed that this pattern was pronounced among men relative to women. This finding aligns with work showing that men endorse traditional gender stereotypes more than women (e.g., Glick & Fiske, 1996) and has implications for how men and women use facial characteristics when forming impressions. For example, men's selections of more masculine faces described by agency as compared to communion could reflect a non-verbal route by which men maintain and reinforce gender roles. Future work may examine this possibility.

Although Experiment 1 supported expectations of gendered facial features based on the big two, several issues were necessary to address. First, like prior work (Walker & Wänke, 2017; Wen et al., 2020), Experiment 1 used few exemplar faces. We thus cannot rule out that the emergent pattern was unique to exemplars rather than reflecting broader expectations of faces. Second, Experiment 1 used altered face image continuums from female to male faces. This choice means that we cannot disentangle whether the current pattern reflect expectations of gendered facial features versus category-based expectations. Experiment 2a addressed these issues.

6. Experiment 2a

In Experiment 2a, we replicated and extended Experiment 1 by manipulating sexually dimorphic facial features *within-gender* to determine whether expectations of gendered facial features occur *within-gender*. Prior work supports that within-gender expectations emerge

(e.g., Walker & Wänke, 2017). For example, people rate more masculinized relative to more feminized faces (regardless of face gender) as being more agentic and less communal (Wen et al., 2020). These patterns suggest sexually dimorphic expectations of faces rather than expectations based on gender categories. If such findings extend to expectations of gendered facial features, people should select masculinized (versus feminized) faces for male and female targets described by agentic relative to communal traits. Notably, gender categories could also *accentuate* expectations of gendered facial features between genders based on the big two. If true, we would expect agentic traits to yield the most expectations of masculinized features when faces are of men relative to women. Likewise, communal traits should yield the least expectations of masculinized features when faces are of men relative to women.

Manipulating gendered facial features within-gender also allowed us to explore the perceiver gender effects from Experiment 1. For example, it could be that men are more likely than women to expect masculinized men's than women's faces specifically when endorsing agentic traits. Men could also be more likely than women to expect feminized women's than men's faces specifically when endorsing communal traits. Indeed, men may believe that associating with femininity threatens their masculinity (Moss-Racusin, Phelan & Rudman, 2010; Stanaland, Gaitner & Gassman-Pines, 2022). These patterns could again reflect a non-verbal route to gender role reinforcement. Because traits commonly used in tasks evaluating agency (e.g., competent) and communion (e.g., warm) relatively positive, however, it could also be that men endorse more masculine male than female faces across trait types so that masculine men are most positively evaluated. Such a pattern would complement work showing that men endorse and encourage high masculinity among men to perpetuate their high status (Vandello & Bosson, 2013).

6.1. Method

6.1.1. Participants

We recruited 225 participants from MTurk and who did not complete Experiment 1 for Experiment 2a. Twelve were excluded for failing an attention check ("Using a mouse click, indicate whether you looked at: Faces, Cars"), yielding an analyzed sample of 213 participants ($M_{age} = 40.96$ years, $SD = 13.9$; $M_{years\ of\ education} = 15.68$, $SD = 2.59$; 85 identifying as female). The same attention check was used for all the experiments. Of the 213 participants, 179 identified as White, 14 as Black, 16 as Asian, two as multi-racial, one as American Indian/Alaska Native, and one self-reported their race as being unknown. Of these participants, 197 identified as non-Hispanic.

6.1.2. Stimuli

We selected 30 male and 30 female White neutrally expressive faces from the Chicago Face Database (Ma et al., 2015). Experiments 2a, 2b, 3a, and 3b all used the same set of 30 men's and 30 women's faces. Using database norms, we verified that the male and female faces had similar age, $t(58) = 0.84$, $p = .40$, $d = 0.21$, 95% CI [-0.29, 0.72], and attractiveness, $t(58) = 0.41$, $p = .68$, $d = 0.11$, 95% CI [-0.40, 0.61]. Thus, any effects of face gender resulting from manipulating the faces' gendered cues could not be attributed to base differences in the faces (e.g., an attractiveness halo). See the Supplemental Material for exact norm information and thus the distributions of these norms.

We used FaceGen Modeller Core 3.14, a software widely used in face perception work (e.g., Cassidy & Liebenow, 2021) to manipulate gendered facial cues (Fig. 3a). FaceGen algorithms yield the relative gender reflected by each face on a scale with five tick marks along a slider scale (Very Masculine, Masculine, Neutral, Feminine, and Very Feminine). These tab sliders were derived from statistical analysis of human faces. In FaceGen, sliders manipulating features such as shape and color are based on linear projections into FaceGen's "face space," which consists of 50 dimensions of symmetric shape, 30 dimensions of asymmetric shape and 50 dimensions of symmetric color. Relevant here,

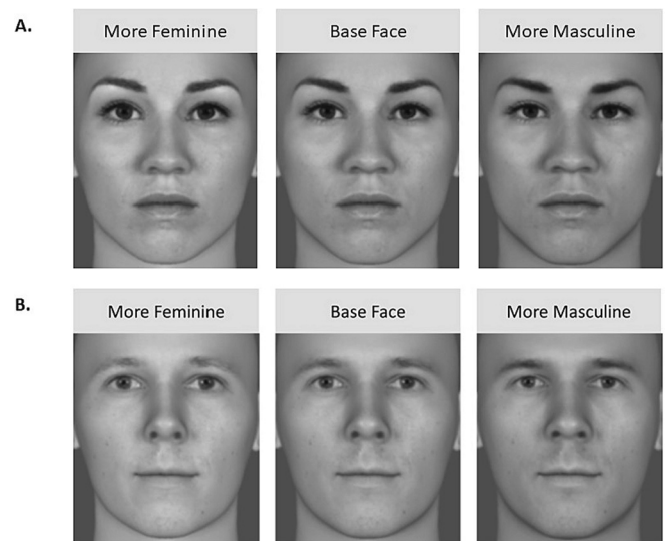


Fig. 3. Example female (A) and male (B) faces manipulated to be more feminine (left) and more masculine (right) in Experiment 2a.

the gender slider is a linear regression on their data set in that same "face space." The gender slider is derived from the difference between the male and female models (see Supplemental Material). The gender slider is made independent of racial group by taking the gender differences only within racial groups and averaging them. It is also made independent of age by removing the projection onto the age slider as described in the FaceGen manual (see Supplemental Material). Using this slider thus reflects a face-valid manipulation of gendered facial features. Using these tick marks, we created a more masculine and a more feminine version of each face. This terminology is consistent with the FaceGen gender slider. We note that this procedure means that the faces were altered on an overall gender continuum, suggesting the manipulation of actual sexual dimorphic differences among faces along one dimension.

That gender was manipulated on a single scale in FaceGen supports the sexual dimorphism framework used in the current work and in much related work (Little, Burriss, Jones & Roberts, 2007; Mitteroecker, Windhager, Müller & Schaefer, 2015; Perrett et al., 1998; Sutherland et al., 2015; Wen et al., 2020). Further, database norms for masculinity and femininity (Ma et al., 2015) for the target faces were strongly linked ($r(58) = -0.97$, $p < .001$), suggesting the appropriateness using one continuum. However, we acknowledge recent work showing that masculinity and femininity can be treated as distinct and orthogonal dimensions of person perception when perceiving faces (Hester, Jones & Hehman, 2021). A discussion of these constructs as separable is beyond the scope of the current work.

To create a masculinized face, we shifted the face's gender placement the length of one full tick mark toward masculinity relative to where the face's gender was set by FaceGen. To create a feminized face, we shifted the face's gender placement the length of one full tick mark toward femininity relative to where the face's gender was set by FaceGen. For example, if a face's gender placement was halfway between Neutral and Feminine, we shifted the gender placement halfway between Feminine and Very Feminine to make a feminized version of the face. This procedure resulted in 60 masculinized and 60 feminized versions of the face identities. To parallel Experiment 1, we used Adobe Photoshop to grayscale and apply a 30% blur to all resulting images and cropped them below the chin and at the top of the forehead.

6.1.3. Procedure

People completed a task comprised of two randomly presented blocks of 30 trials each. The blocks differed by whether participants made choices about faces reflecting agency (i.e., dominance) or

communion (i.e., warmth). Participants evaluated 30 face pairs in each block (for a total of 60 face pairs). Face pair presentation within each block was randomized. Before each block, participants read, “In this section, you will make decisions about a series of face pairs. Your job is to decide which of the two faces looks more dominant [warm]. A dominant [warm] person can also be described as assertive, independent, or ambitious [caring, friendly, and compassionate]. Press ‘S’ to select the face on the left. Press ‘K’ to select the face on the right. The task will advance when you make a decision. Move quickly through the task, but remember it is important that you try your best and go with your gut impression. There are no right or wrong responses. You will be making decisions about 30 face pairs.”

After reading instructions, people completed attention checks verifying they understood the definitions of dominance [warmth]. The question read, “You will be deciding which face is more dominant [warm]. What is another way to describe a dominant [warm] person? (Funny, **Assertive**, Easy-going, Confused [**Caring**, Mean, Annoying, Controlling]).” Twelve were excluded for at least one incorrect response.

Each task trial included a question (“Who is more dominant [warm]?”) above one face pair. Each face pair comprised the masculinized and feminized altered face image of the same identity. Face pair presentation was randomized within-blocks. Each identity was seen once during the task. We made this methodological choice to ensure that initial decisions about an identity could not potentially affect subsequent decisions about the same identity. The more masculine altered face image appeared on the right and left side of the screen approximately half of the time. Four task versions counterbalanced whether people made dominance or warmth decisions about each face pair and the side of the screen the masculinized altered face image of each pair appeared.

After the task, people indicated the extent to which they paid attention and followed task instructions using a scale ranging from 1 (not at all) to 7 (completely) ($M = 6.92$, $SD = 0.31$). They also indicated whether they looked at faces, cars, music, or schools. All analyzed participants indicated faces.

6.1.4. Participant characterization

Participants completed the measures from Experiment 1 in a random order. Responses to political ideology items (Cronbach’s $\alpha = 0.96$) were averaged to create a composite score ($M = 4.65$, $SD = 1.81$). Items measuring hostile (Cronbach’s $\alpha = 0.94$; $M = 1.68$, $SD = 1.29$) and benevolent (Cronbach’s $\alpha = 0.88$; $M = 1.89$, $SD = 1.08$) sexism were averaged to make composite scores. SRQ responses (Cronbach’s $\alpha = 0.88$) were averaged to make a social role belief score ($M = 34.25$, $SD = 22.47$).

6.2. Results

6.2.1. Gender differences emerged on the participant characterization measures

Gender difference patterns for hostile sexism, benevolent sexism, traditional social role beliefs, political ideology and years of education paralleled Experiment 1. See Table 1b for descriptive and inferential statistics.

6.2.2. The big two elicited expectations of gendered facial features

6.2.2.1. Analytic plan. In a mixed effects model, face image selection (masculinized = 1, feminized = 0) was logistically regressed on Trait Type (−1 = dominant and 1 = warm), Face Gender (−1 = man and 1 = woman), Perceiver Gender (−1 = man and 1 = woman) and their interactions as fixed effects. See Table 2a for all coefficient information. The random effects structure included by-participant random intercepts and random slopes for Trait and Face Gender, and by-face identity random intercepts and random slopes for Trait Type (see Table 3a). Note that including this random effects structure acknowledges and accounts for any variance among the face identities. That is, even if any differences among the faces emerged along any metric, that variance is accounted for in the model. Thus, emergent effects (e.g., the expected Trait Type effect) should not be attributed to differences in the faces beyond the gender manipulation.

6.2.2.2. Expectations of facial masculinity. Like Experiment 1, the Trait Type effect showed that people selected the masculinized face as being more likely to be dominant (Estimate = 0.86, 95% CI [0.83, 0.98]) than warm (Estimate = 0.14, 95% CI [0.12, 0.16]), $b = -1.82$, $SE = 0.10$ $t = 18.29$, $p < .001$, 95% CI [−2.01, −1.64]. As expected, the Trait Type effect was qualified by Face Gender, $b = 0.19$, $SE = 0.06$ $t = 2.99$, $p = .003$, 95% CI [0.07, 0.29]. When selecting a face expected to be more dominant, people were less likely to select the masculinized altered face image of a female (Estimate = 0.83, 95% CI [0.79, 0.87]) than a male (Estimate = 0.88, 95% CI [0.85, 0.91]) face, $OR = 0.67$, $SE = 0.12$, $z = 2.17$, $p = .03$, 95% CI [0.47, 0.96]. When selecting the face expected to be more warm, people were more likely to select the masculinized face of a female (Estimate = 0.16, 95% CI [0.14, 0.19]) than a male face (Estimate = 0.12, 95% CI [0.10, 0.15]), $OR = 1.45$, $SE = 0.17$, $z = 3.14$, $p = .002$, 95% CI [1.15, 1.82]. A sensitivity analysis for this interaction effect indicated a minimum detectable effect of log odds = 0.175 with power = 0.80 and alpha = 0.05.

An interaction between Face Gender and Perceiver Gender also emerged, $b = 0.08$, $SE = 0.03$ $t = 3.10$, $p = .002$, 95% CI [0.03, 0.13]. Men were marginally less likely to select a masculinized female (Estimate = 0.49, 95% CI [0.46, 0.52]) than male (Estimate = 0.53, 95% CI

Table 2
Regression summaries for Experiments 2a, 2b, 3a, and 3b.

	a. Experiment 2a			b. Experiment 2b			c. Experiment 3a			d. Experiment 3b		
	Log Odds (SE)	z	p	Log Odds (SE)	z	p	Log Odds (SE)	z	p	Log Odds (SE)	z	p
Trait Type	−1.82 (0.10)	−18.28	< 0.001	−0.25 (0.09)	−2.82	0.0005	−0.82 (0.06)	−12.82	< 0.001	−0.68 (0.06)	−10.78	< 0.001
Face Gender	−0.01 (0.04)	−0.16	0.87	−0.13 (0.07)	−1.96	0.05	0.02 (0.04)	0.49	0.62	−0.005 (0.05)	−0.10	0.92
Participant Gender	−0.05 (0.04)	−1.27	0.20	0.17 (0.08)	2.25	0.02	−0.06 (0.06)	−0.95	0.34	0.07 (0.06)	1.08	0.28
Trait Type × Face Gender	0.19 (0.06)	2.99	0.003	0.03 (0.06)	0.44	0.66	0.12 (0.04)	3.38	0.001	0.10 (0.03)	3.40	0.001
Trait Type × Participant Gender	−0.11 (0.08)	1.36	0.17	0.002 (0.07)	0.03	0.97	0.03 (0.06)	0.50	0.62	−0.17 (0.06)	−2.83	0.005
Face Gender × Participant Gender	0.08 (0.03)	3.10	0.002	0.01 (0.03)	0.39	0.70	−0.02 (0.02)	−1.02	0.31	−0.03 (0.03)	−1.18	0.24
Trait Type × Face Gender × Participant Gender	0.04 (0.03)	1.60	0.11	−0.03 (0.02)	−1.11	0.27	−0.003 (0.02)	−0.15	0.88	−0.001 (0.02)	−0.06	0.95

Table 3

Random effects statistics (SD [95% CI]) for Experiments 2a, 2b, 3a, and 3b.

	a. Experiment 2a	b. Experiment 2b	c. Experiment 3a	d. Experiment 3b
Participants	0.34 [0.27, 0.42]	0.99 [0.88, 1.12]	0.79 [0.70, 0.89]	0.87 [0.88, 1.12]
Participants: Trait Type	1.08 [0.97, 1.22]	0.91 [0.81, 1.03]	0.77 [0.69, 0.88]	0.80 [0.81, 1.03]
Participants: Face Gender	0.01 [0.01, 0.08]	0.15 [0.07, 0.23]	0.14 [0.05, 0.22]	0.22 [0.07, 0.23]
Participants: Trait Type $\times$ Face Gender	0.07 [0.02, 0.14]	0.10 [0.03, 0.16]	0.12 [0.27, 0.20]	0.05 [0.03, 0.16]
Face Identity	0.23 [0.17, 0.30]	0.48 [0.39, 0.60]	0.25 [0.19, 0.32]	0.33 [0.28, 0.47]
Face Identity: Trait Type	0.43 [0.36, 0.53]	0.41 [0.33, 0.51]	0.20 [0.14, 0.26]	0.14[0.33, 0.51]

[0.50, 0.57]) face when endorsing traits, $OR = 0.84$, $SE = 0.08$, $z = 1.87$, $p = .06$, 95% CI [0.71, 1.01]. Women did not exhibit a preference ($Estimate_{female\ face} = 0.51$, 95% CI [0.47, 0.55]; $Estimate_{male\ face} = 0.47$, 95% CI [0.43, 0.51]), $OR = 1.15$, $SE = 0.12$, $z = 1.36$, $p = .18$, 95% CI [0.94, 1.42]. Breaking down this interaction in another way, men were more likely than women to endorse masculinized male faces across traits, $OR = 1.28$, $SE = 0.12$, $z = 2.71$, $p = .01$. By contrast, men and women were similarly endorsed masculinized female faces across traits, $OR = 0.94$, $SE = 0.079$, $z = -0.771$, $p = .44$.

6.3. Discussion

People expected targets described by agentic relative to communal traits to have masculinized facial features. This pattern extended Experiment 1 in several ways. First, Experiment 2a had more face identities, suggesting this effect generalizes across faces rather emerges via idiosyncrasies within a few exemplars. Second, this effect emerged *within* face gender, suggesting sexually dimorphic expectations of faces rather than only category-based expectations. Third, by finding that face gender qualifies the trait effect on expectations, we show that visual gender categories (viewing male or female faces) *accentuate* expectations of sexually dimorphic facial features. Indeed, agentic traits elicited the most expectations of masculinized features when faces depicted men relative to women. Likewise, communal traits elicited the least expectations of masculinized features when faces depicted men relative to women. These findings complement work showing different valenced face impressions of sexually dimorphic facial features depending on target face gender (Oh et al., 2019; Sutherland et al., 2015; Wen et al., 2020). Expectations of sexually dimorphic facial features based on the big two are thus constrained to some extent by target gender.

An open question regards whether the described effects reflect differences in perceiving gender and gender stereotypes above and beyond effects of sexually dimorphic facial characteristics. Future work may address this possibility using faces with a wider range of sexually dimorphic features coupled with faces rated on their masculinity and femininity as well as inferences of gender stereotypic traits. Although beyond the scope of the current work, it would be useful for future work to determine the relations between these unique but interrelated constructs.

Experiment 2a allowed for more exploration of perceiver gender effects. Here, men were more likely to select masculinized male than female faces across trait types. This finding might seem surprising because one might expect men (relative to women) to expect masculinized male relative to female faces specifically when agentic traits describe them. Although the traits used in Experiment 2a are commonly used agentic and communal traits (Abele, 2003; Diehl et al., 2004; Rosette & Tost, 2010); however, they are relatively positive masculine and feminine traits (Diekman & Eagly, 2000). One possibility is that men endorsed more masculine male than female faces across traits so masculine men would be most positively evaluated. Such a pattern would complement work showing that men endorse and encourage men's high masculinity (Vandello & Bosson, 2013). If that is the case, men should not endorse masculinized male faces described by negative traits. Experiment 2b addressed this possibility.

7. Experiment 2b

Experiment 2b extended Experiment 2a by using negative agentic and communal traits. This change provided two benefits. First, we could test for expectations of gendered facial features when targets were described by negative agentic versus communal traits and if face gender again qualified this effect. Such patterns would suggest effect a generalizable expectation effect. Second, we reasoned that if men selected masculinized male than female faces to preserve positivity associated with men's masculinity, then this pattern should not emerge for targets described by negative traits. If anything, this pattern could reverse such that men would be less likely than women to expect masculinized male faces as reflecting negative traits.

7.1. Method

7.1.1. Participants

We recruited 225 participants from MTurk who did not complete the prior experiments. Twenty-six were excluded for failing an attention check, yielding an analyzed sample of 199 participants ($M_{age} = 40.02$ years, $SD = 13.33$; $M_{years\ of\ education} = 15.28$, $SD = 2.15$; 87 identifying as female). Of the analyzed participants, 156 identified as White, 19 as Black, 17 as Asian, six as multi-racial, and one as American Indian/Alaska Native. Of the analyzed participants, 182 identified as non-Hispanic.

7.1.2. Procedure

Experiment 2b replicated Experiment 2a with the following changes. Because the traits from Experiment 2a were positive, we selected "hostile" and "nagging" as two negative traits. We selected these traits from a validated list of gendered traits (Diekman & Eagly, 2000) to complement and contrast the use of "dominant" and "warm" in Experiment 2a. Whereas "dominant" and "hostile" were classified as stereotypically masculine traits, they were classified, respectively, as positive and negative masculine traits. Whereas "warm" and "nagging" were classified as stereotypically feminine traits they were classified, respectively, as positive and negative feminine traits. Participants again evaluated 30 face pairs in each block (for a total of 60 face pairs).

After block instructions, people verified they knew the trait to be evaluated ("Which trait will you be evaluating these faces on?" [silly, easy-going, **hostile**, confused | caring, nice, **nagging**, funny]). Twenty-six people were excluded for failing at least one of these items. People indicated good task adherence ($M = 6.92$, $SD = 0.31$). All indicated that the task involved faces.

7.1.2.1. Participant characterization. Responses to political ideology items were averaged to create a composite score (Cronbach's $\alpha = 0.96$; $M = 4.69$, $SD = 1.72$). Items measuring hostile (Cronbach's $\alpha = 0.94$; $M = 1.46$, $SD = 1.28$) and benevolent (Cronbach's $\alpha = 0.89$; $M = 1.79$, $SD = 1.12$) sexism were averaged to make composite scores. SRQ responses were averaged to make a social role belief score (Cronbach's $\alpha = 0.89$; $M = 26.88$, $SD = 19.39$).

7.2. Results

7.2.1. Gender differences emerged on the participant characterization measures

Gender difference patterns for hostile sexism, benevolent sexism, traditional social role beliefs, political ideology, and years of education paralleled Experiments 1 and 2a. See Table 1c for descriptive and inferential statistics.

7.2.2. The big two elicited expectations of gendered facial features

Like Experiment 2a, a Trait Type effect, $b = -0.25$, $SE = 0.09$, $z = 2.82$, $p = .005$, 95% CI $[-0.42, -0.07]$, showed that people selected the masculinized face as being more likely to be hostile (Estimate = 0.76, $SE = 0.03$, 95% CI $[0.70, 0.81]$) than nagging (Estimate = 0.65, $SE = 0.03$, 95% CI $[0.60, 0.70]$).

Unique to Experiment 2b was a Perceiver Gender effect, $b = 0.17$, $SE = 0.07$, $z = 2.25$, $p = .02$, 95% CI $[0.02, 0.32]$. Here, women (Estimate = 0.74, $SE = 0.02$, 95% CI $[0.69, 0.79]$) selected the masculinized face more than men (Estimate = 0.67, $SE = 0.03$, 95% CI $[0.62, 0.72]$). Contrasting Experiment 2a, no interaction between Perceiver Gender and Face Gender emerged, $b = 0.01$, $SE = 0.03$, $z = 0.39$, $p = .70$. A sensitivity analysis for this interactive effect indicated a minimum detectable effect of log odds = 0.19 with power = 0.82 and alpha = 0.05.

There was also a Face Gender effect, $b = -0.13$, $SE = 0.07$, $z = -1.96$, $p = .05$, 95% CI $[-0.27, 0.002]$. People selected the masculinized face more for male (Estimate = 0.73, $SE = 0.02$, 95% CI $[0.68, 0.78]$) than for female (Estimate = 0.68, $SE = 0.02$, 95% CI $[0.63, 0.73]$) faces. See Table 2b for other coefficient information and Table 3b for random effects information.

7.3. Discussion

People expected targets described by *negative* agentic relative to communal traits to appear masculinized. This pattern suggests a face expectation effect emergent across a range of valenced traits reflecting the big two. Unlike Experiment 2a, face gender did not qualify this expectation. Because negative versus positive information more heavily weighs into impressions (e.g., Rozin & Royzman, 2001), this influence of negative traits could outweighs nuance from gender categories in facial feature expectations. Indeed, negative information is more attention-grabbing than positive (Pratto & John, 1991), leading to more impression updating (Shen & Ferguson, 2021). Because morality-related traits weigh especially heavily into impressions (Wojciszke, 2005), it may be beneficial for future work to probe these effects by manipulating whether negative traits reflect morality or competence.

Women had more masculine expectations of facial appearance relative to men regardless of face gender and trait description. This reversal from Experiment 2a aligns with work showing that men endorse masculinity in other men to maintain men being positively viewed (Reigeluth & Addis, 2021; Stanaland et al., 2022; Vandello & Bosson, 2013). Expecting masculine men to have negative traits would be inconsistent with this motivation.

8. Experiment 3a

Experiments 1 and 2a-b showed expectations of gendered facial features based on the traits with which targets are described. An open question regards the consequences for preferences for targets linked to those expectations. Addressing this question can identify whether expectations of gendered facial features based on the big two affects opportunities afforded to others. Experiment 3a addressed this question through decisions about who people expected to win a workplace award given to targets for displaying agentic or communal behaviors.

Focusing on workplace consequences is relevant because deviating from prescriptive stereotypes exacerbates workplace gender disparities (Phelan et al., 2008; Rudman & Glick, 1999, 2001). For example, people

believe that men relative to women are more and less suited, respectively, for stereotypically masculine and feminine jobs (Glick, Wilk & Perreault, 1995). We tested if such findings are reflected in expectations of gendered facial features among people who may excel at tasks that are more agentic or communal. Because people expected targets with masculinized and feminized features to have, respectively, more agentic and communal traits, we predicted people would expect targets described as being nominated an award for agentic relative to communal workplace behaviors to have masculinized features. We expected face gender to *accentuate* these expectations as in Experiment 2a.

Since perceiver gender effects in the other experiments built on an expectation that men want to preserve societal hierarchy, we wanted to confirm that men and women differed in their preference for hierarchy. Participants thus completed a social dominance orientation scale (SDO7; Pratto, Sidanius, Stallworth, & Malle, 1994). Finally, since the perceiver gender effects might also be explained by a sense of threatened gender identity, we also characterized participants' gender identity threat over a single item. Men perceivers could be threatened by their gender identity, which could parallel their differential expectations of gendered facial features.

8.1. Method

8.1.1. Participants

We again recruited 225 people from MTurk who did not complete the prior experiments. Eight were excluded from analyses due to user error (e.g., entering an incorrect survey code) and three were excluded for failing to pass an attention check, yielding an analyzed sample of 214 participants ($M_{age} = 40.29$ years, $SD = 13.17$; $M_{years\ of\ education} = 15.51$, $SD = 2.41$; 101 identifying as female). Of these 214 participants, 167 identified as White, 16 as Black, 22 as Asian, five as multi-racial, two as Native Hawaiian or Pacific Islander, one as American Indian/Alaska Native, and one self-reported their race as being unknown. Of the 214 participants, 194 also identified as non-Hispanic.

8.1.2. Procedure

Experiment 3a replicated Experiment 2a with the following changes. Instead of selecting a face expected to possess a trait, people selected a face they would nominate for a workplace award for displaying agentic or communal traits. The agentic block instructions read, "You must nominate one of your co-workers for an award. The award should go to a person who has shown themselves to be ambitious, determined, and independent. Who would you nominate?" The communal block instructions read, "You must nominate one of your co-workers for an award. The award should go to a person who has shown themselves to be warm, compassionate, and friendly. Who would you nominate?"

After agentic block instructions, people indicated whether they would be evaluating faces, animals, furniture, or events. After communal block instructions, people indicated whether they would be evaluating faces, places, food, or clothes. Three were excluded for incorrect responses. The analyzed participants indicated task adherence ($M = 6.82$, $SD = 0.58$) and all indicated the task involved faces.

8.1.2.1. Participant characterization. Participants completed the measures from the previous experiments and two new measures in a random order. Responses to political ideology items were averaged to create a composite score (Cronbach's $\alpha = 0.95$; $M = 4.80$, $SD = 1.65$). Items measuring hostile (Cronbach's $\alpha = 0.94$; $M = 1.49$, $SD = 1.21$) and benevolent (Cronbach's $\alpha = 0.90$; $M = 1.91$, $SD = 1.12$) sexism were averaged to make composite scores. SRQ items were averaged to make a social role belief score (Cronbach's $\alpha = 0.89$; $M = 28.45$, $SD = 19.19$).

8.1.2.1.1. Social dominance orientation. People completed an eight-item social dominance orientation questionnaire (Ho et al., 2015) to measure their preference for inequality among groups (Pratto et al., 1994) on a scale 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.”) to quantify social dominance orientation (Cronbach’s $\alpha = 0.92$; $M = 2.15$, $SD = 1.39$).

8.1.2.1.2. Identity threat. To measure gender identity threat, people responded to the item: “To what extent (if at all) have you had concerns that you’d be judged negatively based on your gender?” using a scale ranging from 1 (*not at all*) to 5 (*extremely*) ($M = 2.39$, $SD = 1.37$).

8.2. Results

8.2.1. Gender differences emerged on the participant characterization measures

Gender differences for hostile sexism, benevolent sexism, traditional sole role beliefs, political ideology, and years of education replicated those for Experiments 1, 2a, and 2b. Unique to Experiment 3a, women were higher in identity threat than men, and men were higher in social dominance orientation than women. See Table 3a for descriptive and inferential statistics.

8.2.2. The big two elicited expectations of gendered facial features

Replicating Experiments 1, 2a, and 2b, a Trait Type effect emerged, $b = -0.82$, $SE = 0.06$, $z = -12.82$, $p < .001$, 95% CI $[-0.94, -0.69]$. People selected the masculinized altered face image as being more likely to receive the agentic (Estimate = 0.51, $SE = 0.02$, 95% CI $[0.47, 0.55]$) than the communal (Estimate = 0.17, $SE = 0.01$, 95% CI $[0.14, 0.20]$) award.

Replicating Experiment 2a, a Trait Type by Face Gender interaction emerged, $b = 0.12$, $SE = 0.04$, $z = 3.38$, $p < .001$, 95% CI $[0.05, 0.19]$ (Fig. 4b). People were more likely to choose a masculinized male (Estimate = 0.54, $SE = 0.02$, 95% CI $[0.49, 0.58]$) than female (Estimate = 0.49, $SE = 0.03$, 95% CI $[0.44, 0.54]$) face for a workplace award described by agentic traits. People were less likely to choose a

masculinized male (Estimate = 0.15, $SE = 0.02$, 95% CI $[0.12, 0.18]$) than female (Estimate = 0.19, $SE = 0.02$, 95% CI $[0.16, 0.22]$) face for an award described by communal traits. A sensitivity analysis for this focal interaction indicated a minimum detectable effect of log odds = 0.11 with power = 0.82 and alpha = 0.05. See Table 2c for other coefficient information and Table 3c for random effects information.

8.3. Discussion

Replicating and extending Experiments 1 and 2a-b, people expected targets nominated for an award for agentic relative to communal workplace behavior to have masculinized facial features. Expectations of facial masculinity based on the big two thus extend to workplace decisions for which the big two are relevant.

Replicating Experiment 2a, face gender qualified a trait type effect on expectations. Nominations for a more agentic award elicited the most expectations of masculinized features when faces were of men relative to women. Likewise, nominations for a more communal award elicited the least expectations of masculinized features when faces were of men relative to women. Note that our task design meant that people selected one of two men or two women. It was thus not possible for overall gender imbalances to emerge through, for example, selecting more men than women for the agentic award. This design has a key benefit by highlighting that gender roles may be enforced even when gender inequity cannot emerge. Speculatively, people may rely on expectations of sexually dimorphic features to reinforce gender roles when broader gender inequity is prohibited. Future work manipulating between- and within-gender decisions may examine this possibility.

Unlike Experiment 1, 2a, and 2b, expectations did not differ by perceiver gender. This lack of effect did not emerge even though, as in the prior experiments, men scored higher than women on measures of sexism and traditional social role endorsement. Moreover, men had

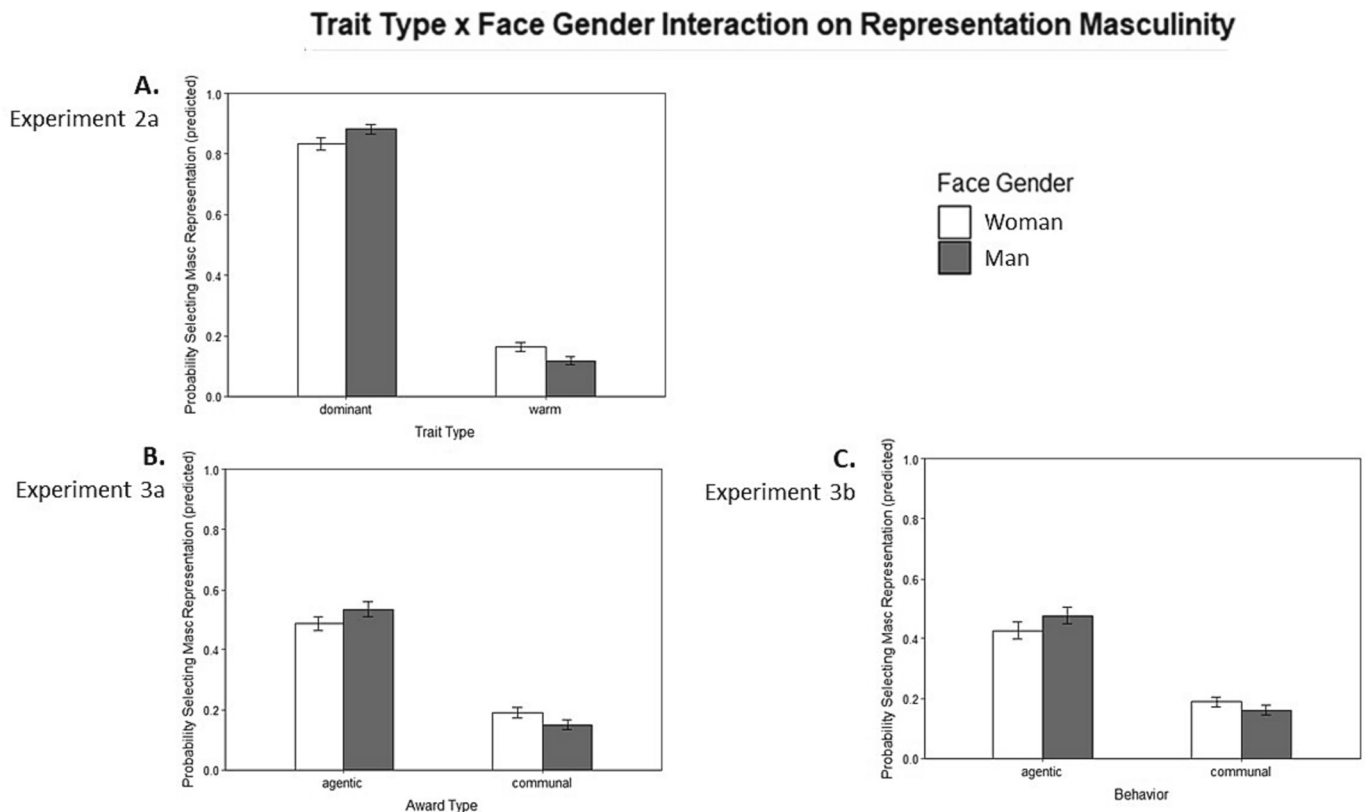


Fig. 4. Interactions between Trait Type and Face Gender emerged in Experiments 2a (A), 3a (B), and 3b (C). Across experiments, participants were more likely to choose more masculine male relative to female faces as reflecting agency. Participants were less likely to choose more masculine male relative to female faces as reflecting communality.

higher social dominance orientation than women, a pattern consistent with work showing that men are more likely than women to uphold traditional social hierarchies (Ho et al., 2015; Pratto et al., 1994; Swami et al., 2013). What, then, might explain this lack of gender differences?

Notably, the prior experiments regarded expectations related to personality traits. By contrast, Experiment 3a regarded expectations contributing to future workplace experiences. In the prior experiments, we reasoned that men relative to women expecting more masculine faces described only by positive traits was consistent with men endorsing masculinity as a positive attribute (Vandello & Bosson, 2013). Overall masculinized expectations that contribute to future workplace experiences, by contrast, may be inconsistent with the social role reinforcement expected from people higher in social dominance orientation and traditional social role beliefs (e.g., Christopher & Wojda, 2008; Radke, Hornsey, Sibley & Barlow, 2018). Speculatively, men may be more willing to endorse men for positive personality traits that happen to be communal, but not nominate them for awards that reinforce behaviors incongruent with prescriptive gender stereotypes. Indeed, men often do not encourage other men to perform tasks countering gender prescriptions (Moss-Racusin, 2014; Moss-Racusin et al., 2010; Reigeluth & Addis, 2021; Stanaland et al., 2022). Future work may examine this possibility.

Our manipulations of the big two have, thus far, explicitly stated traits. We thereby cannot rule out that the expectation effects emergent across experiments simply reflected reactions to these traits. People, however, spontaneously make trait inferences based on behavioral information (e.g., Todorov & Uleman, 2002, 2003, 2004). Such findings suggest expectations of gendered facial features even when traits reflecting the big two are merely inferred. Experiment 3b addressed this possibility.

9. Experiment 3b

Experiment 3b tested whether people have expectations of gendered facial features based on workplace behavioral descriptions that inferred agentic or communal traits. Because people spontaneously infer traits from behavioral information and link them to actors (e.g., Todorov & Uleman, 2003, 2004), we hypothesized that people would expect targets to have masculinized features when described as being more helpful for an analytic, relative to an interpersonal, task. We also expected face gender to accentuate trait effects, as it did in Experiments 2a and 3a.

9.1. Method

9.1.1. Participants

We recruited 225 people from MTurk who did not complete the prior experiments. Five were excluded from analyses due to task error (e.g., entering an incorrect survey code) and seven for failing an attention check, yielding an analyzed sample of 213 participants ($M_{age} = 38.7$ years, $SD = 12.17$; $M_{years\ of\ education} = 15.28$, $SD = 2.39$; 118 identifying as female). Of these 213 participants, 159 participants identified as White, 24 as Black, 21 as Asian, seven as multi-racial, and one self-reported their race as being unknown. Of the 213 participants, 199 identified as non-Hispanic.

9.1.2. Procedure

Experiment 3b replicated Experiment 3a with the following changes. People chose which face they would want to help them on task emphasizing agency or communality. In the agentic block, they read, “You’re stuck on a challenging analytical task at work. Nothing you try seems to be working. Who do you think could solve the problem?” and then made decisions on 30 face pairs. In the communal block, they read, “You are having an issue with a co-worker. Who do you think would be more understanding when discussing this issue?” and then made decisions on a different 30 face pairs. Seven people were excluded for not responding “faces” after reading block instructions. Analyzed

participants indicated task adherence ($M = 6.89$, $SD = 0.45$) and indicated the task involved faces.

9.1.2.1. Participant characterization. Responses to political ideology items were averaged to create a composite score (Cronbach’s $\alpha = 0.97$; $M = 4.61$, $SD = 1.78$). Items measuring hostile (Cronbach’s $\alpha = 0.93$; $M = 1.44$, $SD = 1.16$) and benevolent (Cronbach’s $\alpha = 0.88$; $M = 1.89$, $SD = 1.08$) sexism were averaged to make composite scores. SRQ responses were averaged to make a social role belief score (Cronbach’s $\alpha = 0.88$; $M = 27.13$, $SD = 19.11$). Social dominance orientation items were averaged to make a composite score (Cronbach’s $\alpha = 0.92$; $M = 2.17$, $SD = 1.32$). Participants indicated gender identity threat ($M = 2.52$, $SD = 1.39$).

9.1.2.2. Validating agentic and communal behaviors. To validate that behaviors reflected as agency and communion, people evaluated traits best suited for each behavior using a scale ranging from 1 (*not at all*) to 7 (*extremely*) in two blocks after the characterization measures. In each block, people read the analytical [interpersonal] behavior and evaluated it on the six agentic and six communal traits from Experiment 1 in a random order. Whether people first evaluated the analytical or the interpersonal behavior was randomized. We created composite agentic and communal trait scores for the analytical (Agentic: Cronbach’s $\alpha = 0.85$, $M = 5.38$, $SD = 1.23$; Communal: Cronbach’s $\alpha = 0.92$, $M = 4.35$, $SD = 1.48$) and interpersonal (Agentic: Cronbach’s $\alpha = 0.91$, $M = 3.77$, $SD = 1.48$; Communal: Cronbach’s $\alpha = 0.91$, $M = 5.99$, $SD = 1.03$) behaviors.

Composite trait scores were entered into a 2 (Trait: agentic, communal) $\times$ 2 (Behavior: analytical, interpersonal) repeated-measures ANOVA. Qualifying a main effect of Trait, $F(1,212) = 82.78$, $p < .001$, $\eta_p^2 = 0.28$, was an expected Trait $\times$ Behavior interaction, $F(1, 212) = 295.53$, $p < .001$, $\eta_p^2 = 0.58$. People evaluated the analytical issue as more agentic than communal, $t(212) = 9.24$, $p < .001$, $d = 0.75$. They evaluated the interpersonal issue as more communal than agentic, $t(212) = 18.65$, $p < .001$, $d = 1.74$. There was no effect of Behavior, $F(1, 212) = 0.14$, $p = .71$, $\eta_p^2 < 0.01$.

9.2. Results

9.2.1. Gender differences emerged on the participant characterization measures

Gender differences for hostile sexism, benevolent sexism, traditional social role beliefs, political ideology and years of education paralleled the prior experiments. Women had more identity threat and were older than men. No difference emerged in social dominance orientation. See Table 3b for descriptive and inferential statistics.

9.2.2. The big two elicited expectations of gendered facial features

Replicating Experiments 1, 2a, 2b, and 3a, a Trait Type effect, $b = -0.68$, $SE = 0.06$, $z = -10.78$, $p < .001$, 95% CI $[-0.80, -0.56]$, showed that people selected the masculinized face as being more likely help with the workplace analytical issue (Estimate = 0.51, $SE = 0.02$, 95% CI $[0.47, 0.55]$) than the interpersonal issue (Estimate = 0.17, $SE = 0.01$, 95% CI $[0.14, 0.20]$).

Replicating Experiments 2a and 3a, a Trait Type by Face Gender interaction emerged, $b = 0.10$, $SE = 0.03$, $z = 3.40$, $p < .01$, 95% CI $[0.04, 0.16]$ (Fig. 4c). To help with a workplace analytical issue, people were more likely to choose the co-worker with a masculinized male (Estimate = 0.54, $SE = 0.02$, 95% CI $[0.49, 0.58]$) than female (Estimate = 0.49, $SE = 0.03$, 95% CI $[0.44, 0.54]$) face, $OR = 0.81$, $SE = 0.09$, $z = -1.92$, $p = .05$. To help with a workplace interpersonal issue, people were less likely to choose the co-worker with a masculinized male (Estimate = 0.16, $SE = 0.02$, 95% CI $[0.12, 0.18]$) than female (Estimate = 0.19, $SE = 0.02$, 95% CI $[0.16, 0.22]$) face, although it was not significant, $OR = 1.21$, $SE = 0.15$, $z = 1.56$, $p = .12$. A sensitivity analysis for

this focal interaction indicated a minimum detectable effect of log odds = 0.10 with power = 0.81 and alpha = 0.05.

Unique to Study 3b was a Trait Type by Perceiver Gender interaction, $b = -0.17$, $SE = 0.06$, $z = -2.83$, $p < .01$, 95% CI $[-0.29, -0.05]$. To help with a workplace analytical issue, surprisingly, men (Estimate = 0.39, $SE = 0.03$, 95% CI $[0.33, 0.46]$) selected fewer masculinized faces than women (Estimate = 0.51, $SE = 0.03$, 95% CI $[0.45, 0.58]$), $OR = 0.62$, $SE = 0.12$, $z = -2.58$, $p = .01$. To help with a workplace interpersonal issue, men (Estimate = 0.19, $SE = 0.02$, 95% CI $[0.15, 0.23]$) and women (Estimate = 0.16, $SE = 0.02$, 95% CI $[0.13, 0.20]$) were similarly likely to choose the masculinized face, $OR = 1.22$, $SE = 0.20$, $z = 1.20$, $p = .23$. See Table 2d for other coefficient information and Table 3d for random effects information.

9.3. Discussion

People expected targets who would be helpful for an analytical relative to communal workplace issue to have masculinized features. Replicating Experiments 2a and 3a, face gender qualified this effect. People were more likely to expect a male relative to a female target helpful for an analytical issue to have more masculinized features. Likewise, people were less likely to expect a male relative to a female target helpful with an interpersonal issue to have more masculinized features. The expectation effects from the prior experiments are thus unlikely to simply reflect reactions to stated traits. Indeed, ratings validated that more agentic and communal traits were inferred from, respectively, the analytical and interpersonal workplace issues. Just as trait inferences from behaviors are bound to actors (e.g., Todorov & Uleman, 2002, 2003), these findings suggest that expectations of gendered facial features based on inferences of the big two can be as well. The qualification of the expectation effect by face gender again suggests that expectations based on trait inferences are, in part, constrained by available gender information.

Perceiver gender qualified the trait type effect. Surprisingly, women were more likely than men to expect a masculinized face to help with an agentic issue. By contrast, men and women were similarly likely to expect a masculinized face to help with a communal issue. This pattern contrasted the perceiver gender effects in Experiments 1 and 2a. Although we can only speculate as to why this pattern emerged, one potential explanation may lie in the fact that Experiment 3b was the only experiment in which traits were not explicitly stated. Prescriptive gender stereotypes perpetuate career-related gender stereotypes (e.g., Geis, 1993). Although both men and women explicitly prefer occupations that align with prescriptive stereotypes (Eddleston, Veiga & Powell, 2006), women's implicit preferences align more with these stereotypes than men's do (Gadassi & Gati, 2009). One possibility is thus that, at least for agentic workplace tasks, women's preferences may align more with gender stereotypes than men's when inferences from the big two are inferred. Because women reported higher identity threat than men, another possibility is a heightened salience of threat may have made gender norms more salient as well (Sinclair, Carlsson & Björklund, 2016). Speculatively, this salience could explain why women were more likely than men to select more masculine faces to help with the agentic issue. Future work may examine these possibilities.

10. Experiment 4

In the prior experiments, the face stimuli varied on one dimension. This constrained manipulation allowed us to experimentally detect whether people expect different sexually dimorphic facial features based on agentic and communal trait information. A limitation of this manipulation is that participant endorsements were constrained to faces differing on this dimension. Thus, it is unclear whether endorsements truly reflected people's expectations or whether they emerged as a function of the forced-choice nature of the task. Experiment 4 addressed this limitation using a reverse correlation paradigm to test whether

people have natural expectations of sexually dimorphic facial features based on agency and communality.

Reverse correlation estimates how people mentally represent stimuli (visual, auditory, etc.; see Jack & Schyns, 2017 for an overview). Here, we use reverse correlation to estimate how people mentally represent faces. Reverse correlation allows a way to assess participants' expectations because mental representations of faces can vary in a large variety of ways. By having people evaluate the gendered features of these mental representations, we can determine if, out of the variety of ways mental representations may vary, differences in the gendered nature of faces emerged in mental representations of agentic and communal faces. Because previous social psychology work shows that mental representations can reflect distinct characteristics (e.g., Dotsch & Todorov, 2012), reverse correlation is an ideal method to assess the hypothesis that trait inferences from the "big two" produce gendered expectations of facial features.

10.1. Method

10.1.1. Participants

One hundred fifty people who did not complete the prior experiments were recruited from MTurk. Thirty-one were excluded from analyses due to failing an attention check, yielding an analyzed sample of 119 participants ($M_{age} = 38.98$ years, $SD = 10.79$; $M_{years\ of\ education} = 15.26$, $SD = 2.33$; 44 identifying as female). Of these 119 participants, 93 identified as White, 15 as Black, seven as Asian, and four as multi-racial. Of the 119 participants, 106 identified as non-Hispanic.

10.1.2. Phase 1: Face Classification

In the first phase, people generate classification images (CIs) via a face classification task. A CI is a mental representation of a face quantified by the combined average of all image choices on a given prompt. Here, people were randomly assigned to select which face of a pair appeared more agentic ($N = 60$) or more communal ($N = 59$). Men and women were evenly distributed across these task versions, $X^2(1, N = 119) = 0.25$, $p = .62$.

10.1.2.1. Stimuli. For a base face, we created an average of the neutrally expressive average male and the neutrally expressive average female face from the Karolinska Face Database (Lundqvist, Flykt, & Öhman, 1998). Next, we generated one hundred trials consisting of two faces each. These faces were derived from the base face by overlaying randomly generated noise patterns onto it (for more details, refer to Dotsch & Todorov, 2012). For each trial, the face pair consisted of the base face was combined with a unique noise pattern, and then its reverse counterpart.

10.1.2.2. Task and image processing. Participants completed 100 self-paced trials presented at random. Each trial consisted of a face pair presented side by side. Participants selected the face that appeared more agentic [communal]. Participants then completed the characterization measures as in the other experiments. Responses to political ideology items were averaged to create a composite score (Cronbach's $\alpha = 0.93$; $M = 4.67$, $SD = 1.61$). Items measuring hostile (Cronbach's $\alpha = 0.92$; $M = 1.87$, $SD = 1.18$) and benevolent (Cronbach's $\alpha = 0.88$; $M = 2.37$, $SD = 1.04$) sexism were averaged to make composite scores. SRQ responses were averaged to make a social role belief score (Cronbach's $\alpha = 0.82$; $M = 39.35$, $SD = 20.30$). Social dominance orientation items were averaged to make a composite score (Cronbach's $\alpha = 0.90$; $M = 2.75$, $SD = 1.44$).

To create a CI for each participant, the noise patterns from each participant's 100 selections were averaged and overlaid onto the base face. Each CI represents the participant's unique mental representation of an agentic [communal] face (see Fig. 5 for examples). Whereas some work (van Rijsbergen, Jaworska, Rousselet, & Schyns, 2014) applies



Fig. 5. Examples of agentic (A & B) and communal (C & D) CLIs from Experiment 4.

statistical analysis to, for example, the pixel intensities of CLIs, assess featural differences, much social perception work (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Goh et al., 2021; Sutherland et al., 2015; Wen et al., 2020) is less concerned with specific featural differences than more general trait impressions of the CLIs. We used the latter approach here.

10.1.3. Phase 2: Face ratings

The Phase 1 participants' unique representations of gendered facial features were estimated by the ratings of an independent group of 60 naïve raters who did not complete any prior experiments. One rater was excluded from analyses due to failing an attention check, yielding an analyzed sample of 59 raters ($M_{age} = 40.09$ years, $SD = 15.09$; $M_{years\ of\ education} = 14.68$, $SD = 4.50$; 24 identifying as female). Of these 59 raters, 38 participants identified as White, 17 as Black, two as Asian, and one as American Indian/Alaska Native. Of the 59 raters, 54 identified as non-Hispanic. Responses to political ideology items were averaged to create a composite score (Cronbach's $\alpha = 0.94$; $M = 4.79$, $SD = 1.58$).

The raters saw the 119 unique CLIs from Phase 1 in a random order and were unaware of what prompt yielded the CLIs. Each CI was rated on its gendered facial features on a scale ("How masculine (1) or feminine (7) does this face look?") ranging from 1 (*very masculine*) to 7 (*very feminine*). Note using participant-level CLIs reduces the possibility of Type I error inflation in reverse correlation findings (Cone, Brown-Iannuzzi, Lei, & Dotsch, 2021).

10.2. Results

10.2.1. Gender differences emerged on the participant characterization measures

Among the Phase 1 participants, gender differences for hostile sexism, benevolent sexism, traditional social role beliefs, political ideology and years of education paralleled the prior experiments. No difference emerged in social dominance orientation or age. See Table 1 for descriptive and inferential statistics.

10.2.2. The big two elicited expectations of gendered facial features

Phase 2 Gender Rating was regressed on Phase 1 Trait ($-1 =$ agentic and $1 =$ communal), and Phase 1 Perceiver Gender ($-1 =$ man, $1 =$ woman), and their interaction as fixed effects.

This model included a random effects structure such that intercepts were allowed to vary by participant and by CI. We allowed a Trait Type

effect to vary by participant and CI.

Replicating the previous experiments, a Phase 1 Trait effect emerged, $b = 0.21$, $SE = 0.05$, $t = 4.40$, $p < .001$, 95% CI [0.12, 0.30]. People mentally representing communal faces had more feminine representations (Estimate = 4.40, $SE = 0.13$, 95% CI [4.14, 4.65]) than people mentally representing agentic faces (Estimate = 3.98, $SE = 0.16$, 95% CI [3.67, 4.29]). We confirmed high interrater reliability across CLIs (119 mental representations; $\alpha = 0.97$). We also separately validated high interrater reliability across the agentic (60 mental representations; $\alpha = 0.81$) and communal (59 mental representations; $\alpha = 0.88$) CLIs. There was no Phase 1 Perceiver Gender effect, $b = 0.01$, $SE = 0.04$, $t = 0.36$, $p = .72$, 95% CI [0.12, 0.30], and no interaction, $b = 0.01$, $SE = 0.04$, $t = 0.37$, $p = .71$, 95% CI [0.12, 0.30].

See the Supplemental Material for exploratory analyses. These additional analyses were done to explore potential differences between masculinity and femininity as theoretically separable dimensions, as well as to detect potential categorical differences instead of just featural ones. All results support the premise that people have gendered expectations of facial features based on the big two.

10.3. Discussion

Experiment 4 conceptually replicated the prior experiments while addressing a limitation of Experiments 2–3. Using an open-ended and implicit task where people could represent agentic and communal faces with a large variety of featural combinations, people still had gendered expectations of facial features. That is, people expected communal faces be more feminized than agentic faces. This experiment suggests that expectations of gendered facial features is not only a replicable expectation but is also a robust expectation. Although agentic faces were rated as more masculine than communal faces, that does not mean that these faces only differed by their gendered features. Indeed, although the reverse correlation paradigm is open ended in its generation of mental representations within a face space, it is not open ended to the extent that experimenters choose how these representations are subsequently evaluated. Future work may explore different ways in which agentic and communal representations of faces may vary (e.g., in attractiveness) to explore how distinct gendered representations of agentic and communal faces are.

Participant gender did not qualify this expectation effect or affect it overall. Although we can only speculate on why this might be, it could be that men are more motivated to have high evaluations on positive traits in explicitly assessed tasks. In tasks assessing more implicitly assessing expectations, both men and women may similarly have expectations aligning with gendered traits. Indeed, men and women have prescriptive stereotypes with men expected to possess agentic traits and women expected to possess communal traits (Bruckmüller & Abele, 2013; Martin & Slepian, 2020; Wiggins, 1991). Future work may directly address this possibility. We also note that unlike the previous experiments, men and women did not vary on any of the participant characterization measures. It could also be that gender differences in expectations of gendered facial features may only emerge in samples endorsing difference attitudes and beliefs about gender.

In the present work, the base face used is a gender-ambiguous morph of a man and woman's face that has been previously used in multiple reverse correlation experiments (Cassidy & Krendl, 2018; Degner et al., 2019; Oliveira et al., 2019). It could be fruitful future work to explore how using different base faces in this type of paradigm could influence the present results. From the categorical data in the present study, we do have indirect evidence that there is categorical distinction, not just distinction of sexually dimorphic facial features. However, this is beyond the scope of the current paper. Thus, future work could benefit from exploring if there would be a gender category shift depending on the gender of the base face by using different base faces in a similar paradigm.

By having naïve raters evaluate classification images generated by

participants, we take a holistic approach to assessing mental representations that has been used to answer a wide variety of social psychological research questions (e.g., [Giacomin, Tskhay & Rule, 2022](#); [Gingras et al., 2023](#); [Imhoff et al., 2013](#); [Petsko, Lei, Kunst, Bruneau & Kteily, 2020](#)) and that is detailed on primers on using reverse correlation to examine social perception ([Brinkman, Todorov & Dotsch, 2017](#); [Degner et al., 2019](#); [Dotsch & Todorov, 2012](#); [Dotsch, Wigboldus, Knippenberg & A., 2013](#); [Hess, Hareli & Scarantino, 2023](#); [Imhoff et al., 2013](#); [Oliveira et al., 2019](#); [Oliver, Tracy, Young & Wout, 2023](#)). This approach is well-suited to address whether people generally represent faces as more feminized or masculinized based on trait descriptions. That we used a multi-level model on participant-level CIs that reduces concerns about Type I error ([Cone et al., 2021](#)), and that allowed for random rater-level variability bolsters confidence in the above-described findings.

Although the above-described approach is widely used ([Brinkman et al., 2017](#); [Degner et al., 2019](#); [Dotsch & Todorov, 2012](#); [Giacomin et al., 2022](#); [Hess et al., 2023](#); [Oliveira et al., 2019](#); [Petsko et al., 2020](#)), one limitation of this approach is that it does not formally analyze the CIs themselves to distinguish any statistically significant featural differences of the representations between conditions (see [Jack, Caldara, & Schyns, 2012](#)). We note the objective importance of conducting such formal analyses of CIs, as one aspect of this limitation is that any of the CIs might not include statistically significant features and, therefore, could comprise noise akin to a single trial. This limitation means it might be unlikely that several CIs would be rated in the same way. Because the purpose of Experiment 4 was to determine general expectations of traits reflected in faces and not to derive the significant differing featural content underlying expectations, we did not conduct such formal analyses. However, we confirmed high interrater reliability across all CIs and within agentic and communal CIs. Although high interrater reliability may suggest the presence of significant featural content, future research should determine specific featural detail contributing to the present findings. Such work could also vary the type of traits (e.g., manipulating the number of communal and agentic traits in reverse correlation prompts) to determine how featural content of faces changes based on shifts in a target's relative agency and communion. Thus, this future work would provide insight about how people systematically use facial features to arrive at their evaluations.

11. General discussion

Although some work has suggested a link between gendered trait inferences from faces during person perception ([Oh et al., 2019](#); [Sutherland et al., 2015](#); [Todorov & Uleman, 2002, 2003](#)), the current work is novel in that it consistently shows this link using a variety of contexts and methods, emphasizing the robust generalizability and replicability of this link. The current work provides consistent evidence that people have expectations for facial masculinity and femininity based on descriptions reflecting the big two. Here, people consistently expected more facial masculinity for targets paired with agentic relative to communal traits (Experiments 1, 2a-b) and workplace behaviors (Experiments 3a-b). This expectation effect emerged in both between-gender (Experiment 1) and within-gender (Experiments 2a-b, 3a-b) contexts, regardless of whether traits were explicitly stated (Experiments 1, 2a-b, 3a) or inferred (Experiment 3b), regardless of trait valence, and when expectations were assessed in a more implicit and open-ended way (Experiment 4). In within-gender decision contexts, face gender accentuated trait effects in that more masculine male relative to female faces were consistently expected more for agentic traits and workplace tasks, but consistently expected less for communal traits and workplace tasks (Experiments 2a, 3a-b). Finally, we provide exploratory analyses showing that expectation effects may differentially vary by perceiver gender across contexts. These experiments extend past literature on facial expectations and representations based on gendered concepts (e.g., [Degner et al., 2019](#); [Imhoff et al., 2013](#)) by showing, in a variety of contexts, that people not only have facial expectations based

on traits and behaviors that directly relate to sexually dimorphic facial features. These findings will be discussed in turn.

Across experiments, people expected faces associated with agentic relative to communal traits and workplace behaviors to be masculinized. This effect parallels work on prescriptive stereotypes (e.g., [Prentice & Carranza, 2002](#)) and the big two ([Abele, 2003](#); [Bruckmüller & Abele, 2013](#)). Moreover, this finding forges a link between work showing agentic and communal trait inferences in *reaction to*, respectively, sexually dimorphic facial features ([Cassidy & Liebenow, 2021](#); [Walker & Wänke, 2017](#); [Wen et al., 2020](#)) to the *expectations* reflecting work prescriptive gender stereotypes. People not only react to sexually dimorphic facial features in ways that reflect the big two, but they expect it. Extending theoretical work asserting that gender is a fundamental lens by which people experience the world ([Martin & Slepian, 2020](#)), it could be that gender is so engrained in our society that it affects perception by influencing what people expect from faces. Indeed, people expect teachers in STEM (relative to language) to appear more masculine ([Degner et al., 2019](#)). Notably, recent work suggests that the facial characteristics of leaders convey expectations people have of STEM environments ([Joshi, Lloyd, Diekmann & Hugenberg, 2022](#)). Considered with the expectation effects from the current work, these findings suggest that sexually dimorphic facial features may serve as cues that reinforce traditional gender roles in workplace environments. Indeed, even when gender inequity was prohibited when people chose between two male or two female faces in Experiments 2a-b and 3a-b, faces with specific gendered features were expected to align with descriptions. Some men and women may still not be considered for tasks reflecting the big two based on their facial features. This possibility has broader implications for increasing minoritized groups' representation in myriad domains by highlighting the importance of within-category cues in understanding underrepresentation.

In within-gender contexts, face gender qualified the overall expectation effect when descriptions were positive (Experiments 2a, 3a-b), but not negative (Experiment 2b). Gender categories may thus accentuate expectations of sexually dimorphic facial features. Put another way, expectations of sexually dimorphic facial features based on the big two may be, to some extent, constrained by target gender. A male relative to a female category cue may enhance expectations of high masculinity given agentic descriptions, but reduce such expectations given communal descriptions. These findings again complement work on prescriptive gender stereotypes using primarily verbal cues (e.g., [Prentice & Carranza, 2002](#)) by showing that although sexually dimorphic facial features exert strong influence on person perception (e.g., [Sutherland et al., 2015](#)), gender category information nevertheless plays a vital role in how people reason about the big two.

The current findings have implications for research on backlash discrimination by providing a new lens by which to consider this phenomenon. Theoretical and empirical work on backlash discrimination, for example, asserts that deviating from communal prescriptions yields backlash against women aspiring to leadership roles characterized by agency ([Brescoll, Okimoto & Vial, 2018](#); [Okimoto & Brescoll, 2010](#); [Rudman & Glick, 2001](#)). In the current work, we show that people are more likely to expect a more masculine face for traits and roles characterized by their agency. However, we do not consider other trait inferences of these faces. In a situation where people select between a more masculine male and female face for help on a communal workplace task, for example, that people are more likely to select the male face. Consistent with backlash research (e.g., [Rudman & Glick, 2001](#)), potential insufficient communality conveyed by a masculine female face may yield backlash against that target for a feminized task. Indeed, people have more negative reactions toward counter-stereotypic female than male faces ([Sutherland et al., 2015](#)). Future research may consider this possibility.

Relatedly, it will be important to evaluate people's *expected satisfaction* of an agentic task performed by a target with a masculine male and female face. Consistent with role congruity theory ([Eagly & Karau,](#)

2002), it could be that people expect to be less satisfied by behaviors of a target with a masculine female relative to male face due to deviations from gender stereotypes. It could also be that people expect to be more satisfied by agentic behaviors performed by a female target with feminized versus masculinized facial features because such a target would be perceived as maintaining communal prescriptions. Indeed, people more likely to support Kamala Harris's vice-presidential candidacy in 2020 perceived her face as more feminized (Cassidy & Liebenow, 2021).

Across experiments, we show that perceiver gender may differentially affect expectations of sexually dimorphic facial features. Although we stress that these findings are exploratory, we provide some speculation on them for the benefit of future research. In Experiments 1 and 2a, men were more likely than women to endorse masculinized male faces across traits, a pattern we interpreted from the lens of work showing that men are more likely to endorse and positively evaluate masculinity (Vandello & Bosson, 2013). Experiment 2b corroborated this possibility by showing that women are more likely to expect masculinized features than men across negative traits. Interestingly, men were no longer more likely to endorse masculinized features of male faces than women when considering workplace behaviors (Experiment 3a and 3b). In fact, women were more likely to have expectations of faces in line with prescriptive gender stereotypes when traits were inferred. We offer a potential explanation rooted in literature showing that relative to men, women's implicit, but not explicit, job preferences reflect gender stereotypes to a greater extent (Gadassi & Gati, 2009). Coupled with our consistent finding that men scored higher than women on measures of sexism, social role endorsement, and social dominance orientation, these findings suggest that perceiver gender effects on expectations of sexually dimorphic facial features are quite nuanced. Because gender is proposed to be a fundamental lens by which people view the world (Martin & Slepian, 2020), it could be that women continue to be affected by that lens in their expectations of sexually dimorphic facial features given the automaticity of evaluating facial cues (Todorov, 2008) and binding traits to actors (e.g., Todorov & Uleman, 2002) despite having lower explicitly measured sexism, social role endorsement, and social dominance orientation than men.

Facial masculinity and femininity can be treated as distinct and potentially orthogonal dimensions of person perception in that they interact to explain unique variation within traits (Hester et al., 2021). Approaching gendered facial features from the lens of sexual dimorphism may thus not explain all differences in expectations based on the big two. Future work may consider manipulating faces across multiple dimensions to assess unique expectations of facial androgyny, masculinity, and femininity, across traits. Beyond strengthening our understanding of the link between face perception and prescriptive gender stereotypes, this work may better characterize how people have expectations of targets who may not "fit" prescriptive gender stereotypes. For example, future researchers may examine how perceived sexuality may complicate the present findings. Indeed, people who appear gender-atypical are more likely to be perceived as gay or lesbian (Rule, Ambady, Adams, & Macrae, 2008; Rule, Macrae, & Ambady, 2009; Rule & Alaei, 2016), which could complicate inferences from the big two.

The current experiments used younger White faces to assess expectation effects. Relevant for future work is research showing that race is gendered (Galinsky, Hall & Cuddy, 2013). For example, Black faces are perceived as more masculine than White and East Asian faces, whereas East Asian faces are perceived as more feminine than Black and White faces (Johnson, Freeman & Pauker, 2012). These associations suggest that expectation effects shown in the current work may be complicated by face race as well as face gender. For example, in comparison to male and female White faces, expectation effects for agentic traits may be blunted among male and female Black faces given that people associated being Black with masculinity. The current work may serve as an important foundation for intersectionality research in backlash, face perception, and person perception.

The current work provides evidence that people have expectations of

sexually dimorphic facial features based on traits and workplace descriptions reflecting the big two. Prescriptive gender stereotypes oft-examined verbally using perceiver self-report (e.g., Prentice & Carranza, 2002) also seem to emerge via a non-verbal route in face perception. Thus, facial features beyond people's control may be one route by which traditional gender roles are reinforced despite people's efforts to obtain gender parity in the workplace. These findings reflect a growing need to consider gendered facial features (e.g., Joshi et al., 2022) when considering interventions that will be most effective in combating gender inequity in a myriad of domains oft-characterized as being agentic or communal.

Author notes

The authors received no external funding for this work. The authors declare no conflict of interest.

Declaration of Competing Interest

There are no conflicts of interest in conducting research or publishing the present work. None of the authors on this manuscript have any conflicts of interest, financially nor personally, that could be viewed as inappropriately influencing our present work.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jesp.2023.104585>.

References

- Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, 85(4), 768–776. <https://doi.org/10.1037/0022-3514.85.4.768>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2), 256–274. <https://doi.org/10.1037/0033-2909.111.2.256>
- Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem sex-role inventory. *Sex Roles: A Journal of Research*, 43(7–8), 499–528. <https://doi.org/10.1023/A:1007119516728>
- Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles*, 54(7–8), 459–467. <https://doi.org/10.1007/s11199-006-9018-y>
- 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>
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162. <https://doi.org/10.1037/h0036215>
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88(4), 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>
- Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS One*, 3(8), Article e3022. <https://doi.org/10.1371/journal.pone.0003022>
- Bonneau, C. W., & Kanthak, K. (2020). Stronger together: Political ambition and the presentation of women running for office. *Politics, Groups, and Identities*, 8(3), 576–594. <https://doi.org/10.1080/21565503.2018.1528159>
- Bosak, J., & Sczesny, S. (2011). Exploring the dynamics of incongruent beliefs about women and leaders: Exploring the dynamics of incongruent beliefs. *British Journal of Management*, 22(2), 254–269. <https://doi.org/10.1111/j.1467-8551.2010.00731.x>
- Brescoll, V. L., Okimoto, T. G., & Vial, A. C. (2018). You've come a long way...maybe: How moral emotions trigger backlash against women leaders. *Journal of Social Issues*, 74(1), 144–164. <https://doi.org/10.1111/josi.12261>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brown, E. R., Thoman, D. B., Smith, J. L., & Diekmann, A. B. (2015). Closing the communal gap: The importance of communal affordances in science career motivation. *Journal of Applied Social Psychology*, 45(12), 662–673. <https://doi.org/10.1111/jasp.12327>

- 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>
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692. <https://doi.org/10.1037/1076-8971.5.3.665>
- Cassidy, B. S., & Krendl, A. C. (2018). Believing is seeing: Arbitrary stigma labels affect the visual representation of faces. *Social Cognition*, 36(4), 381–410. <https://doi.org/10.1521/soco.2018.36.4.381>
- 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.12243. <https://doi.org/10.1111/asap.12243>
- Christopher, A. N., & Wojda, M. R. (2008). Social dominance orientation, right-wing authoritarianism, sexism, and prejudice toward women in the workforce. *Psychology of Women Quarterly*, 32(1), 65–73. <https://doi.org/10.1111/j.1471-6402.2007.00407.x>
- Cole, S., Trope, Y., & Balcetis, E. (2016). In the eye of the betrothed: Perceptual downgrading of attractive alternative romantic partners. *Personality and Social Psychology Bulletin*, 42(7), 879–892. <https://doi.org/10.1177/0146167216646546>
- Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science*, 12(5), 760–768. <https://doi.org/10.1177/1948550620938616>
- Costrich, N., Feinstein, J., Kidder, L., Marecek, J., & Pascale, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology*, 11(6), 520–530. [https://doi.org/10.1016/0022-1031\(75\)90003-7](https://doi.org/10.1016/0022-1031(75)90003-7)
- Davison, H., & Burke, M. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, 56, 225–248. <https://doi.org/10.1006/jvbe.1999.1711>
- Degner, J., Mangels, J., & Zander, L. (2019). Visualizing gendered representations of male and female teachers using a reverse correlation paradigm. *Social Psychology*, 50, 233–251. <https://doi.org/10.1027/1864-9335/a000382>
- Diehl, M., Owen, S. K., & Youngblade, L. M. (2004). Agency and communion attributes in adults' spontaneous self-representations. *International Journal of Behavioral Development*, 28(1), 1–15. <https://doi.org/10.1080/016502503440000226>
- Diekmann, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, 26(10), 1171–1188. <https://doi.org/10.1177/0146167200262001>
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3, 562–571. <https://doi.org/10.1177/1948550611430272>
- Dotsch, R., Wigboldus, D. H. J., Knippenberg, V. A. N., & A. (2013). Behavioral information biases the expected facial appearance of members of novel groups: Biased facial appearance of novel outgroups. *European Journal of Social Psychology*, 43(1), 116–125. <https://doi.org/10.1002/ejsp.1928>
- Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 283–308.
- 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>
- Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A Meta-analytic perspective. *Personality and Social Psychology Bulletin*, 17(3), 306–315. <https://doi.org/10.1177/0146167291173011>
- Eagly, A. H., & Wood, W. (2012). Social role theory. In I. P. Van, A. K. Lange, & E. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). SAGE Publications Ltd.. <https://doi.org/10.4135/9781446249222.n49>
- Eddleston, K. A., Veiga, J. F., & Powell, G. N. (2006). Explaining sex differences in managerial career satisfaction preferences: The role of gender self-schema. *Journal of Applied Psychology*, 91, 437–445. <https://doi.org/10.1037/0021-9010.91.2.437>
- Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self-recognition. *Personality and Social Psychology Bulletin*, 34(9), 1159–1170. <https://doi.org/10.1177/0146167208318601>
- 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(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Gadassi, R., & Gati, I. (2009). The effect of gender stereotypes on explicit and implicit career preferences. *The Counseling Psychologist*, 37(6), 902–922. <https://doi.org/10.1177/001100009334093>
- Galinsky, A. D., Hall, E. V., & Cuddy, A. J. C. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, 24(4), 498–506. <https://doi.org/10.1177/0956797612457783>
- Geis, F. L. (1993). Self-fulfilling prophecies: A social psychological view of gender. In *The psychology of gender* (pp. 9–54). Guilford Press.
- Gervais, S. J., & Hillard, A. L. (2011). A role congruity perspective on prejudice toward Hillary Clinton and Sarah Palin: Prejudice toward Clinton and 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. *The Leadership Quarterly*, 33(6), Article 101578. <https://doi.org/10.1016/j.leaqua.2021.101578>
- Gingras, F., Fiset, D., Plouffe-Demers, M.-P., Deschênes, A., Cormier, S., Forget, H., & Blais, C. (2023). Pain in the eye of the beholder: Variations in pain visual representations as a function of face ethnicity and culture. *British Journal of Psychology*, 114(3), 621–637. <https://doi.org/10.1111/bjop.12641>
- 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., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, 79(5), 763–775. <https://doi.org/10.1037/0022-3514.79.5.763>
- 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. (2021). Narrow prototypes and neglected victims: Understanding perceptions of sexual harassment. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspi0000260>
- 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>
- Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, 41(8), 1123–1134. <https://doi.org/10.1177/0146167215591495>
- Hess, U., Hareli, S., & Scarantino, A. (2023). What is it about your face that tells me what you want from me? Emotional appeals are associated with specific mental images. *Cognition and Emotion*, 1–10. <https://doi.org/10.1080/02699931.2023.2266991>
- 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. S., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., ... Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO_r scale. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspi0000033>
- 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>
- Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, 141(1), 19–25. <https://doi.org/10.1037/a0023463>
- Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, 68, 269–297.
- Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, 148(6), 1008–1021. <https://doi.org/10.1037/xge0000591>
- 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., Diekmann, A. B., & Hugenberg, K. (2022). In the face of opportunities: Facial structures of scientists shape expectations of STEM environments. *Personality and Social Psychology Bulletin*, 01461672221077801. <https://doi.org/10.1177/01461672221077801>
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01086>
- 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., Buurkner, P., & Herve, M. (2018). Package “Emmeans”. R Package Version 4.0-3. <http://cran.r-project.org/package=emmeans>
- Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, 28(1), 18–27. <https://doi.org/10.1016/j.evolhumbehav.2006.09.002>
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces—KDEF (CD ROM). Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- Mackie, D. M., Hamilton, D. L., Susskind, J., & Rosselli, F. (1996). Social psychological foundations of stereotype formation. *Stereotypes and Stereotyping*, 41–78.
- 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>
- Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The Morphometrics of “masculinity” in human faces. *PLoS One*, 10(2), Article e0118374. <https://doi.org/10.1371/journal.pone.0118374>
- Moss-Racusin, C. A. (2014). Male backlash: Penalties for men who violate gender stereotypes. *Gender in Organizations*, 247–269.
- 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>
- 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., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 49(5), 888–902. <https://doi.org/10.1002/ejsp.2569>
- Oliver, A., Tracy, R. E., Young, S. G., & Wout, D. A. (2023). Black + white = prototypically black: Visualizing Black and White People's mental representations of black-white biracial people. *Personality and Social Psychology Bulletin*, 01461672231164026. <https://doi.org/10.1177/01461672231164026>
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., ... Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, 394(6696). <https://doi.org/10.1038/29772>. Article 6696.
- Perrine, R. M. (1998). Students' views of the characteristics of Instructors' approachability. *Psychological Reports*, 82(2), 519–525. <https://doi.org/10.2466/pr0.1998.82.2.519>
- Petsko, C., Lei, R., Kunst, J., Bruneau, E., & Kteily, N. (2020). Blatant dehumanization in the mind's eye: Prevalent even among those who explicitly reject it? *Journal of Experimental Psychology: General*, 150. <https://doi.org/10.1037/xge0000961>
- 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>
- Pleasant, R. K. (2011). Men learning feminism: Protecting privileges through discourses of resistance. *Men and Masculinities*, 14(2), 230–250. <https://doi.org/10.1177/1097184X11407048>
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61, 380–391. <https://doi.org/10.1037/0022-3514.61.3.380>
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741–776. <https://dash.harvard.edu/handle/1/3207711>
- 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/10.1111/1471-6402.101-1-00066>
- Radke, H. R. M., Hornsey, M. J., Sibley, C. G., & Barlow, F. K. (2018). Negotiating the hierarchy: Social dominance orientation among women is associated with the endorsement of benevolent sexism. *Australian Journal of Psychology*, 70(2), 158–166. <https://doi.org/10.1111/ajpy.12176>
- Reigeluth, C. S., & Addis, M. E. (2021). Policing of Masculinity Scale (POMS) and pressures boys experience to prove and defend their "manhood." *Psychology of Men & Masculinities*, 22, 306–320. <https://doi.org/10.1037/men0000318>
- van Rijsbergen, N., Jaworska, K., Rousselet, G. A., & Schyns, P. G. (2014). With age comes representational wisdom in social signals. *Current Biology*, 24(23), 2792–2796. <https://doi.org/10.1016/j.cub.2014.09.075>
- 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>
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2
- 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. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4), 743–762. <https://doi.org/10.1111/0022-4537.00239>
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, 48(1), 165–179. <https://doi.org/10.1016/j.jesp.2011.10.008>
- Rule, N. O., & Alaei, R. (2016). Gaydar": The Perception of Sexual Orientation From Subtle Cues. *Current Directions in Psychological Science*, 25(6), 444–448. <https://doi.org/10.1177/0963721416664403>
- Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology*, 95, 1019–1028.
- Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, 104(3), 409–426. <https://doi.org/10.1037/a0031050>
- Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science*, 20(4), 441–443. <https://doi.org/10.1111/j.1467-9280.2009.02314.x>
- 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, Article 104219. <https://doi.org/10.1016/j.jesp.2021.104219>
- Sinclair, S., Carlsson, R., & Björklund, F. (2016). Getting along or ahead: Effects of gender identity threat on communal and agentic self-presentations. *Scandinavian Journal of Psychology*, 57(5), 427–432. <https://doi.org/10.1111/sjop.12310>
- Stanaland, A., Gaither, S., & Gassman-Pines, A. (2022). When is masculinity "fragile"? An expectancy-discrepancy-threat model of masculine identity. *PsyArXiv*. <https://doi.org/10.31234/osf.io/fgbk9>
- 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>
- Swami, V., Neofytou, R.-V., Jablonska, J., Thirlwell, H., Taylor, D., & McCreary, D. R. (2013). Social dominance orientation predicts drive for muscularity among British men. *Body Image*, 10(4), 653–656. <https://doi.org/10.1016/j.bodyim.2013.07.007>
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83, 1051–1065. <https://doi.org/10.1037/0022-3514.83.5.1051>
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39, 549–562. [https://doi.org/10.1016/S0022-1031\(03\)00059-3](https://doi.org/10.1016/S0022-1031(03)00059-3)
- Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, 87, 482–493. <https://doi.org/10.1037/0022-3514.87.4.482>
- Trapnell, P. D., & Paulhus, D. L. (2012). Agentic and communal values: Their scope and measurement. *Journal of Personality Assessment*, 94(1), 39–52. <https://doi.org/10.1080/00223891.2011.627968>
- Vandello, J., & Bosson, J. (2013). Hard won and easily lost: A review and synthesis of theory and research on precarious manhood. *Psychology of Men & Masculinity*, 14, 101. <https://doi.org/10.1037/a0029826>
- 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), Article 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). <https://doi.org/10.1177/1474704920980642>, 1474704920980642.
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In *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.
- Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, 16(1), 155–188. <https://doi.org/10.1080/10463280500229619>
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, 2(3), 1497–1517. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>