LIEBENOW, HAYLEY. Ph.D. Expectations Of Women: Trait Inferences, Nonverbal Cues, And Their Impact On Women's Underrepresentation In Leadership (2024).
Directed by Dr. Brittany Cassidy. 259 pp.

Women encounter a variety of obstacles when striving for leadership over the course of their careers. These obstacles can take the form of gendered trait expectations of how women and leaders "should" behave – expectations that often clash with one another. Trait inferences from facial cues (often out of women's control) have also been shown to affect evaluations and impressions of women when striving for leadership. Appearance expectations of what a good leader "should" look like may also clash with women's appearance and femininity in general. By identifying factors that contribute to people's impressions and evaluations of women based on differing sexually dimorphic facial features, the current program of research adds to our understanding of how women can hopefully successfully navigate these obstacles. Indeed, work from this dissertation also reveals how nonverbal cues influence perceptions of women and highlight the factors that majorly impact job selections. The present work hopefully illuminates how women are impacted from these cues throughout their careers as they strive for positions of authority.

EXPECTATIONS OF WOMEN: TRAIT INFERENCES, NONVERBAL CUES, AND THEIR IMPACT ON WOMEN'S REPRESENTATION IN LEADERSHIP

by

Hayley Liebenow

A Dissertation
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro

2024

Approved by

_____

Dr. Brittany Cassidy
Committee Chair

DEDICATION

To my dad – thank you for raising me to be the woman I am today. Thank you for teaching me to be bold, to stand tall, and to never doubt myself. I carry your strength and love in every decision, every struggle, and every choice I make. Words will never be enough.

To my sister, Brittany, thank you for telling me what I needed to hear – even when I really didn't want to hear it. Across the continent, know that I'll miss these years when you were so nearby.

To my childhood friends from Georgia and my Wellesley besties, I am who I am today because of you all. Graduate school would have been infinitely harder without your love, support, spontaneous phone calls, and Facetime hangouts to remind me what life is really about.

To my Greensboro friends, who saw me at my highest highs and lowest lows of graduate school, thank you for always being there to pull me back up again.

To my advisor, Brittany Cassidy, thank you for dealing with me and my endless questions for four years. I am the scholar you see today because of you. Without your guidance, I wouldn't know the first thing about proper academic writing or how to run a multi-level model – let alone be writing this dissertation today. To my committee – thank you for the months and months of time, effort, and feedback to get this dissertation to what you see now. I appreciate you all.

And finally, to my younger self, who chose to go through this challenging life experience. We did it!

"It is literally impossible to be a woman…Like, we have to always be extraordinary, but somehow we're always doing it wrong…You have to be a boss, but you can't be mean. You have to lead, but you can't squash other people's ideas. You're supposed to love being a mother, but don't talk about your kids all the damn time. You have to be a career woman, but also always be looking out for other people."

- *America Ferrera, "Barbie" (film)*

APPROVAL PAGE

This dissertation written by Hayley Liebenow has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Brittany Cassidy

Committee Members

Dr. Michael Kane

Dr. Ethan Zell

Dr. Paul Silvia

May 21, 2024

Date of Acceptance by Committee

May 21, 2024

Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTEGRATED INTRODUCTION

Relative to men, women remain conspicuously underrepresented in leadership roles across various sectors, from politics to STEM to academia (Baskaran & Hessami, 2018). This pervasive gender gap has spurred extensive research efforts aimed at comprehending the factors contributing to women's limited presence in positions of authority. Existing literature has extensively delved into theories such as the glass ceiling metaphor (i.e., invisible barriers hindering the advancement of certain demographic groups in the workplace; Eagly & Sczesny, 2009), role congruity (i.e., societal expectations influence the perceived suitability of individuals for leadership roles based on gender stereotypes; Eagly & Karau, 2002), and stereotype threat (i.e., the risk of confirming negative stereotypes, leading to diminished performance by people in situations where those stereotypes are salient; Steiner et al., 2022). Relatively underexplored in this work is systematic tests of how nonverbal cues contribute to this underrepresentation. This absence is surprising because nonverbal cues have long been known to affect impressions of others and social judgements (Hoss et al., 2005; Kleider-Offutt et al., 2021; Lick & Johnson, 2018; D. Martin & Macrae, 2007; Xie et al., 2019).

Nonverbal cues are defined as facial and body indicators that infer information about a person to others (e.g., Carbon, 2011; Knutson, 1996; D. Martin & Macrae, 2007). Relevant to this proposal are inferences of the "big two" traits of social perception – agency and communion (Bruckmüller & Abele, 2013; Hsu et al., 2021; A. E. Martin & Slepian, 2020). Agency encompasses traits including assertiveness, independence, and strength, traditionally aligned with men and masculinity. Communion includes traits such as warmth, empathy, compassion, often stereotypically linked to women and femininity. Inferences of the big two are thus stereotypically gendered in that they can be inferred based on someone’s gender group (Bruckmüller & Abele,

2013; Liebenow et al., 2024; A. E. Martin & Slepian, 2020). Beyond being gendered, trait inferences of agency and communion play a crucial role in shaping individuals' evaluations of leadership potential (Koenig et al., 2011). Critically, the extent of these inferences are associated with systemic biases against women aspiring to leadership (Eagly & Sczesny, 2009; Koenig et al., 2011; Rosette & Tost, 2010a). Women who exhibit high agency may face backlash due to incongruence with societal expectations, challenging traditional gender norms (Bosak & Sczesny, 2011; Eagly & Karau, 2002a; Koenig et al., 2011). Conversely, women who emphasize communion may confront the "double-bind" dilemma, where they are perceived as lacking the assertiveness necessary for leadership roles (Haines & Stroessner, 2019a). This dynamic balance of agency and communion creates a challenging landscape for women aspiring to leadership positions, emphasizing the present need for a more nuanced understanding of how these dimensions impact leadership evaluations and opportunities for women in various contexts over the course of their careers.

Agentic and communal trait inferences can come from nonverbal spaces (e.g., Wen et al., 2020). That these inferences are relatively spontaneous (Willis & Todorov, 2006) suggest that women may be perceived as agentic or communal despite their actual traits. Indeed, while consensual, trait inferences from faces are often inaccurate (D. Martin & Macrae, 2007). In the present program of work, I extend the literature on women's underrepresentation in leadership by examining it through the lens of nonverbal behavior. By examining evaluations of real-life and hypothetical women navigating leadership aspirations, I illustrate that gendered trait inferences from nonverbal cues contribute perceptions of aspiring women leaders. By better understanding perceptions of women in workplace environments, we can gain a deeper understanding of the

barriers women face when striving for leadership positions and potentially discover a new route for interventions to rectify women's underrepresentation in leadership.

In this dissertation, I present three manuscripts examining how gendered trait inferences affect women's leadership prospects. In the first manuscript, I investigate how gendered trait inferences affected an example of real woman – Kamala Harris – striving for an exclusive leadership position, the U.S. Vice President, during the 2020 election (Liebenow et al., 2022). Here, I find validation of past work on the "big two" and women's leadership by demonstrating a strong positive role of and evaluations of Harris regardless of whether her communal or agentic qualities are highlighted. By focusing on a real-life exemplar who is also a woman of color, I also enhance our understanding of how gendered trait inferences affect aspiring women leaders with intersectional social identities.

In the second manuscript, I examine the extent to which people expect targets with agentic or communal qualities to have facial characteristics that appear gendered (i.e., masculinized or feminized) (Liebenow et al., 2024). Here, I show that people have robust expectations that gendered facial features are associated with possessing more agency or communion, and that decisions within the workplace align with these expectations. These findings have broad implications for aspiring women leaders because I demonstrate evidence that people have expectations of sexually dimorphic facial features based on traits and workplace descriptions reflecting the "big two." Thus, facial features beyond women's control may be one route by which traditional gender roles are reinforced in the workplace.

The third manuscript addresses potential consequences of gendered trait inferences from masculinized and feminized facial features (Liebenow et al., under review). Here, I show that aspiring women, but not men, leaders are penalized for possessing masculinized facial features

associated with stronger inferences of agency. The findings from this manuscript suggest that sexually dimorphic facial features, defined as the differing levels of masculinity and femininity in one's facial features, may comprise a nonverbal barrier to aspiring women leaders.

Before presenting this body of work, I detail the literature motivating it. First, I cover existing literature on the "big two" traits of social perception and how these inferences are reflected in gender stereotypes and expectations for women, as well as consequences for women aspiring to leadership roles. Next, I delve into how trait impressions of faces may also affect women's journey to leadership by focusing on inferences of agency and communion. Finally, I discuss the central aims and takeaways of the program of research comprising this dissertation,

Gender Stereotypes and the "Big Two"

Gender stereotypes can manifest in two forms: descriptive stereotypes, which mirror the qualities attributed to men and women (Koenig, 2018), and prescriptive stereotypes, which encapsulate societal expectations regarding what men and women should be (Prentice & Carranza, 2002a). These stereotypes play a vital role in reinforcing traditional social roles (Eagly & Wood, 2012) and beliefs about appropriate behaviors for men and women contribute to differences in how they express agentic and communal behaviors (Eagly & Wood, 1991). Initial research (Bem, 1974) identified traits characterizing societal perceptions of men and women, and now, 50 years later, the present program of literature demonstrates the enduring nature of these gender stereotype-associated traits.

A consistent theme in this area of research is the alignment of emergent traits with the "big two" traits in social cognition (Bruckmüller & Abele, 2013a; A. E. Martin & Slepian, 2020a; Wiggins, 1991). The belief that men should possess traits associated with agency and women with communality is a prevalent pattern amongst these papers' findings. Despite a few

studies suggesting changing perceptions of gender stereotypes (Diekman & Eagly, 2000; Eagly et al., 2020a), a summary of 73 years of public opinion polls indicated persistent gender stereotypes reflecting the "big two." However, recent work still demonstrates how gender prescriptions remain entrenched in American society (Cassidy & Liebenow, 2021a; Liebenow et al., 2024; A. E. Martin & Slepian, 2020a). Indeed, one of the empirical papers within this dissertation (Liebenow et al., 2024) consistently found, across six experiments spanning various social and workplace contexts, that people expect women and faces with highly feminine features to possess communal attributes compared to men and faces with highly masculine features.

Given the linkage between gender and the "big two," it is unsurprising that people who deviate from gender prescriptions can be subject to severe negative consequences. Extensive research has focused on the consequences for women, with role congruity theory suggesting unfavorable evaluations for women who strive for leadership roles (Eagly & Karau, 2002a). Indeed, this unfavorable assessment of female leaders compared to male leaders has been shown in prior work (Heilman et al., 1995), where researchers portrayed highly successful female managers and participants evaluated these women as exhibiting more hostility (i.e., bitter, selfish, etc.) and less rationality (i.e., less logical, adept at distinguishing emotions from thoughts, etc.) compared to successful male managers. Agentic behaviors in women, which fulfill leadership expectations, have been found to be less favorable than agentic behaviors in men – making it challenging for women who are aiming for these leadership roles. For example, women described with agentic traits are considered less suitable for managerial roles than women described with communal traits (Phelan et al., 2008). Agentic female leaders who are *also* seen as communal, however, tend to receive favorable evaluations (Rosette & Tost, 2010).

Empirical evidence and real-world examples underscore this importance of perceived communality for agentic women, as seen in the positive evaluations of Hillary Clinton primarily based on her perceived communality despite being perceived as more agentic (Gervais & Hillard, 2011), supporting a role prioritization perspective. Role prioritization suggests that balancing agentic and communal behaviors prevents negative evaluations, offering a crucial framework for understanding gendered trait inferences in evaluating women aspiring for leadership in male-dominated environments. Men also face negative consequences when deviating from gender prescriptions, such as being perceived as less qualified than women for female-stereotyped jobs, such as nurses and teachers (Davison & Burke, 2000). However, men do not seem to broadly incur penalties for possessing communal traits, and thus a potential agentic deficit. Rather, they seem to incur backlash for perceived low status (Moss-Racusin, 2014; Moss-Racusin et al., 2010). Further understanding the dynamics of gendered trait inferences and their differing impact on men's and women's leadership journeys is pivotal in addressing the challenges women encounter in these male-dominated positions. Through work like the present dissertation, I underscore the importance of better understanding the impact gendered trait inferences and nonverbal cues have on women's leadership prospects in a variety of fields.

Research on gender stereotypes and their connection to the "big two" has primarily utilized verbal measures (Eagly et al., 2020a), with individuals rating the desirability of trait characteristics for men and women on a scale (Fiske et al., 2007; Prentice & Carranza, 2002a; L. A. Rudman et al., 2012). Albeit influential, this work has somewhat overlooked a vast literature showing that people spontaneously and efficiently make consensual trait inferences from facial features (Carbon, 2011; Knutson, 1996; D. Martin & Macrae, 2007). Impressions formed from facial features have a lasting impact on people's perceptions of others (Ambady & Rosenthal,

1992) and often take precedence over other incoming information (Jaeger et al., 2019). From this knowledge, impressions made from a woman's appearance may have a lasting impact on how others view her in the workplace. These gendered impressions from facial cues are important to study, as facial cues outside of women's control could potentially impede their prospects at leadership positions.

Facial Features Elicit Impressions of the "Big Two"

Differing levels of masculinity and femininity in facial features have been used to define the extent of a target person's sexually dimorphic facial features. Masculinized facial features are defined as facial features traditionally associated with masculinity and male faces, such as being wider, longer, and more rectangular than women (Ferrario et al., 1993; Samal et al., 2007), while feminized facial features are facial features traditionally associated with femininity and female faces, such as a smaller nose length and larger eye height (Villavisanis et al., 2023). Specifically, people often use features of eyebrow thickness, width of nose at its base, width of mouth, eye-to-eyebrow distance, forehead height, and distance between inner corners of the eyebrows to make accurate distinctions of gender identification (Burton et al., 1993).

The extent of sexually dimorphic facial features affects people's impressions of others (Dunkle & Francis, 1990; Hughes & Bremme, 2011; Liebenow et al., 2024; Perrett et al., 1998; Walker & Wänke, 2017a; Yang et al., 2015). For example, facial masculinity positively relates to agency-related competence stereotypes (Oh et al., 2019). These associations between facial features and the "big two" have been demonstrated through various methods. Explicit trait rating tasks, such as rating faces varying in masculinity and femininity on 60 different traits (Walker & Wänke, 2017), show the clear associations between sexually dimorphic facial features and the "big two." Implicit measures, such as utilizing a classic seven-step Implicit Association Task

(IAT) reveal people's associations of masculine and feminine faces with, respectively, agentic and communal traits (Wen et al., 2020), indicate the automatic nature of these gendered associations with sexually dimorphic facial features.

Women with feminized facial features and men with masculinized facial features appear *gender-typical* (Hoss et al., 2005). That some people appear gender-typical means that others must appear gender-*atypical* (e.g., women with masculinized features). Notably, people tend to evaluate more gender-typical faces more positively than less gender-typical faces (Oh et al., 2019; Sutherland et al., 2015). Indeed, one prominent study found that only women with gender-atypical (i.e., more masculine or dominant) faces incurred significant bias compared to facially stereotypical men or women's faces (Sutherland et al., 2015). For atypical men's faces, there was no significant difference in negative evaluations. These negative consequences for gender-atypical women's faces seem to mirror the negative outcomes incurred by women deviating from stereotypical gendered trait prescriptions. Because inferences from gender-atypical faces deviate from prescriptive gender stereotypes, it raises the possibility that these trait inferences create a nonverbal barrier perpetuating a status quo of women's underrepresentation leadership roles (e.g., Garcia-Retamero & López-Zafra, 2006). Indeed, work on role congruity theory (Eagly & Karau, 2002) suggests that the perceived mismatch between the communal female gender role and agentic leadership qualities results in negative evaluations of women leaders compared to men (Banchefsky et al., 2016; Gervais & Hillard, 2011). Supporting this general premise, people who visualized Kamala Harris's face as appearing more masculine (versus feminine) were less likely to support her in 2020 (Cassidy & Liebenow, 2021). Further supporting this idea is work showing that the facial masculinity of women politicians negatively relates to the number of votes they receive (Hehman et al., 2014). Therefore, people may have favorably evaluated

Kamala Harris partially due her being portrayed in a more feminine way, thus identifying a nonverbal route to her leadership success.

Because face impressions are consensual but not necessarily accurate (Rule et al., 2013), it raises the possibility that trait inferences of the "big two" from gendered facial features may restrict opportunities for women even if those inferences are ultimately untrue. For instance, predefined expectations about the appearance of employees in relation to their assigned roles may lead to a higher number of job offers for individuals whose facial features align with these stereotypical expectations. Indeed, prior work has shown that people have generally consensual gendered perceptions of job occupations, with one study demonstrating that people perceive some jobs as more masculine and others more feminine (Glick et al., 1995). Another study found evidence that women were perceived as more well-suited for occupations characterized by high warmth and low competence (He et al., 2019). These results may translate into consequences for women who strive toward leadership, as leadership is perceived as masculine (Koenig et al., 2011) and therefore not suitable for women.

Aims of Present Research

In this dissertation, I present three manuscripts contributing to the literature on women's underrepresentation in leadership by examining evaluations of real-life and hypothetical women navigating leadership opportunities. In the first manuscript, I focused on characterizing intersecting effects of agency, communion, and political ideology on evaluations of Kamala Harris during the 2020 U.S. Presidential election. Being perceived as prioritizing communality appears key for women political candidates to receive support (Conroy et al., 2020a). Little work (e.g., Gervais & Hillard, 2011) has examined this theoretical premise in actual politicians, and before now, no work had examined it using a woman candidate of color. This gap is important to

fill because relative to White women, women of color face additional challenges in attaining leadership roles (Nelson & Piatak, 2021). I found that people perceived Harris as more agentic than communal even when her communality was emphasized. This paper extended the literature on gendered evaluations of real female candidates (e.g., Gervais & Hillard, 2011) by showing separable *and* interactive influences of perceived communality and perceiver ideology on evaluations of Harris. Positive communality effects on evaluations of Harris were weaker for more liberal people when a description of conveyed details about her communality (but not her agency). These findings suggest that female candidates cannot simply assume unwavering support from people sharing their political ideology. Instead, I identify in this paper potential sources of bias from gendered trait inferences that can negatively affect female candidates in particular. This work illustrates a need for gendered trait inferences to be more thoroughly considered in campaign strategies, especially for candidates who belong to minoritized groups.

In the second manuscript, I examined whether gendered trait inferences elicit *expectations* of how faces "should" appear. This contrasts prior work that has primarily focused on *reactions* to faces varying in sexually dimorphic features. My contrasting approach is theoretically important for research on women's underrepresentation in leadership because gendered expectations of facial appearance may influence beliefs about which gender group is best suited for certain behaviors and tasks in the workplace. From this approach, we can better understand how job suitability based on nonverbal cues might harm women's chances for leadership roles. 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). 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). I then conceptually replicated expectation effects by showing that mental representations of agentic and communal faces appear correspondingly gendered (Experiment 4). These findings illustrate a nonverbal route by which people make decisions about women's suitability for certain roles based on gender stereotypes that have wide-ranging implications for workplace behavior.

In the third manuscript, I examined whether the extent of sexually dimorphic facial features constitute an obstacle to aspiring women leaders. My approach is theoretically important for research on women's underrepresentation in leadership because I examine the underexplored possibility that facial features may elicit negative outcomes for women early in their careers, connecting perspectives from the "big two" and backlash literatures. To date, research has examined outcomes for aspiring women leaders varying in facial features without the inclusion of aspiring men leaders (e.g., Cassidy & Liebenow, 2021). Other research has shown only gender-atypical women (versus men) faces to be negatively evaluated (Sutherland et al., 2015), suggesting potential backlash against these women. Without evaluative outcomes, it is unclear whether findings from backlash research (Moss-Racusin, 2014; Rudman & Glick, 1999; Rudman & Phelan, 2008) apply to women based on their sexually dimorphic facial features. My work filled this gap in the literature. Manipulated masculinized versus feminized facial features elicited, respectively, higher dominance and lower warmth impressions. Aspiring women leaders with masculinized versus feminized facial features received less favorable evaluations for several leadership roles, whereas men's evaluations were unaffected by varying features. These findings provide novel evidence that sexually dimorphic facial features constitute a barrier unique to aspiring women leaders.

CHAPTER II: UNDERSTANDING EVALUATIONS OF KAMALA HARRIS IN 2020: POLITICAL IDEOLOGY QUALIFIES PERCEIVED COMMUNALITY EFFECTS WHEN COMMUNAL CUES ARE PRESENT

Abstract

Women of color (versus White women) are underrepresented in the United States government. Identifying factors that affect evaluations of these women is important to understand their underrepresentation. Deviating from communal expectations contributes to backlash against women. Being perceived as prioritizing communality thus appears key for women to receive support. Little work, however, has examined this relation in actual politicians and how perceiver political ideology may affect it. We examined how gendered trait inferences and political ideology affected evaluations of Kamala Harris, the first woman of color elected to the executive branch, before the 2020 election. People perceived Harris as more agentic than communal (Studies 1-2). Communal trait inferences and having a more liberal political ideology each positively related to evaluations of Harris. More liberal relative to more conservative perceivers had weaker positive communality effects when evaluating her expected success (Studies 1-2) and when a description conveyed Harris's communality (versus agency; Study 2). These findings highlight communality effects on evaluations of Harris and suggest a context under which she was likely more supported by co-partisans. Moreover, these studies identify potential sources of bias toward female candidates of color, illustrating a need for gendered trait inferences to be thoroughly considered in campaign strategies.

Introduction

Decades of research have converged on the idea that agency and communion are the "Big Two" dimensions of social cognition (Abele et al., 2008; Bem, 1981; Bruckmüller & Abele, 2013a; A. E. Martin & Slepian, 2020a; Wiggins, 1991). Illustrating the reach of these dimensions, agency and communion impressions largely capture how people perceive (Eagly, 1997), evaluate (Phelan et al., 2008), and behave toward (Archer, 1996; Eagly & Wood, 1991) others. These traits have distinct associated inferences. Whereas agency is associated with independence and competence, communion is associated with warmth and a general social orientation (Abele & Wojciszke, 2007; Wiggins, 1991). The social construct of gender is theorized to underlie these inferences (Abele, 2003; A. E. Martin & Slepian, 2020a). People associate agentic and communal traits with men and women, respectively (Eagly & Crowley, 1986; Eagly & Wood, 1991). One consequence of the theorized link between gender and the Big Two is difficulty in disentangling them, which likely perpetuates women's underrepresentation in leadership (e.g., in politics; Okimoto & Brescoll, 2010). Indeed, women often face backlash for agentic behaviors enabling men's success because they are perceived as deviating from communal expectations (e.g., Okimoto & Brescoll, 2010; Rudman & Phelan, 2008). Being perceived as communal may thus be essential for agentic women to attain leadership positions (Cassidy & Liebenow, 2021a; Rosette & Tost, 2010b; Sinclair et al., 2016).

Much work examining effects of agentic and communal trait inferences on evaluations of aspiring female leaders has done so using hypothetical women and contexts (e.g., Okimoto & Brescoll, 2010; Rudman & Glick, 2001; Rudman & Phelan, 2008). Although it is important to characterize evaluations of agentic women (Diehl et al., 2004; L. A. Rudman & Glick, 2001a), less work (e.g., Conroy et al., 2020; Gervais & Hillard, 2011) has examined how these inferences

affect evaluations of *actual* agentic women aspiring to leadership. This area of work is important to characterize whether our theoretical understanding maps onto evaluations of real women.

Social psychological research examining actual agentic women aspiring to high leadership roles in the United States government has, to date, focused on White women (e.g., Hillary Clinton; Gervais & Hillard, 2011). The increasing proportion of non-White individuals in the United States (Myers & Levy, 2018), however, highlights a need to identify how agentic and communal inferences relate to evaluations of women of color who aspire to leadership. Women of color are especially underrepresented in government (e.g., Nelson & Piatak, 2021) and are often evaluated differently than White women when displaying agentic behaviors (Livingston et al., 2012; Motro et al., 2021). Such findings call for an examination of a more diverse sample of aspiring female leaders (Gershon & Lavariega Monforti, 2021; Sanchez-Hucles & Davis, 2010; M. C. Schneider & Bos, 2019) to reveal how gendered trait inferences affect support for agentic women evaluated on the basis of both their race and gender.

The 2020 United States presidential election provided an opportunity to fill this gap given Kamala Harris's vice-presidential nomination. Throughout her career, Harris has held male-dominated positions (e.g., prosecutor; Glick et al., 1995), which has likely led to her being perceived as highly agentic. Beyond being one of the first woman nominated to the Vice Presidency, Harris's identity as a Black and Asian woman made her the first woman of color to be nominated. Harris's racial identity was subject to scrutiny throughout her campaign (Dowe, 2020), with media coverage of her candidacy emphasizing both racial and gender identifiers (Cassese et al., 2022). To this end, we first briefly summarize work on gendered trait inferences of aspiring female leaders and discuss how those inferences may be especially relevant to women of color. Because actual female politicians have well-known political ideologies, we next discuss

how perceiver political ideology may qualify effects of gendered trait inferences on evaluations of aspiring female leaders. Finally, we present two studies conducted before the 2020 election that examined inferences of Kamala Harris as agentic and communal and how those inferences affect evaluations of her.

Gendered Trait Inferences of Aspiring Female Leaders and their Consequences

Sixty years of public opinion polls show that expectations about the traits women possess remain pervasive in the United States (Eagly et al., 2020b). Linked with the traditional female gender role (Eagly & Wood, 2012), both men and women perceive women (relative to men) as having more communal traits (e.g., warm) reflecting a relationship orientation (Eagly et al., 2020b; Eagly & Karau, 2002a; Eagly & Wood, 2012). By contrast, people perceive men (relative to women) as having more agentic traits (e.g., assertive) reflecting leadership and dominance. These expectations are prescriptive in that they dictate how people think others should behave (Prentice & Carranza, 2002a). Thus, the societal belief that women are communal is paralleled by the prescription that women *should be* communal.

Stereotypic prescriptions are theorized to maintain a status quo (Hoffman & Hurst, 1990; Jost & Banaji, 1994) of women's underrepresentation in leadership (e.g., Garcia-Retamero & López-Zafra, 2006). Role congruity theory posits that women are less likely to attain leadership positions when role expectations are incongruent with female gender stereotypes (Eagly & Karau, 2002a). Women in mixed-sex dyads, for example, often acquiesce leadership to men only when a task is stereotypically masculine (Ritter & Yoder, 2004). Beyond simply not attaining leadership positions, women violating communal gender prescriptions often receive negative bias because they face a double-bind when aspiring to leadership (Fiske et al., 2018; Phelan et al., 2008; L. A. Rudman & Glick, 1999, 2001a). First, because leadership-related traits (e.g.,

assertive) are associated with men, women must work harder to show they possess them. Second, because such highly agentic women are often perceived as not being communal, they receive backlash for that prescriptive deviation (e.g., Rudman & Glick, 2001).

More negative evaluations of agentic women are prevalent in the political domain (e.g., Schneider et al., 2022). United States executive offices have been male-dominated, which creates obstacles for women aspiring to them (Smith et al., 2007). Indeed, simply aspiring to political office may violate communal expectations of women and attenuate the likelihood of their receiving support (Conroy et al., 2020; Okimoto & Brescoll, 2010). Because this literature has focused on White women, whether these findings extend to women of color is unclear. Considering this generalizability is important because multiple possibilities emerge for how communal inferences may relate to evaluations of female candidates of color. Indeed, past work indicates that women of color may experience a "strategic advantage" or a "double disadvantage" in support due to their intersecting racial and gender identities (Gershon & Lavariega Monforti, 2021).

Work examining race and gender effects on agentic biases toward women shows that people stereotype Black women as highly dominant, angry, and hard-working (Rosette et al., 2016). This work suggests little overlap between these stereotypes of Black women and the communal stereotypes attributed to the superordinate category of women. Thus, one possibility is that lessened communal prescriptions for Black women may result in highly agentic Black women being less dependent on perceived communality to attain leadership roles. These women may not be perceived as violating communal expectations with their high agency. Consistent with this possibility, some work has shown that Black women are not conferred the lower status that White women are when they are highly agentic (Livingston et al., 2012). Although Kamala

Harris is multi-racial, she is often perceived as Black by the electorate (Ma et al., 2021). These perceptions may be attributed to the "one drop rule" in that people who are visibly of Black racial heritage are perceived as fully Black in society (Hollinger, 2005). Indeed, less than 20% of people identify Harris's racial identity as both Black and Asian (Clayton et al., 2022). Evaluations of Harris may thereby parallel evaluations expected toward Black women. Perceived communality may thus only have a weak, if any, positive relation with evaluations of Harris before 2020 election.

Women of color, however, remain especially underrepresented in politics and organizations. They are less likely than White women, for example, to attain leadership roles in government (Nelson & Piatak, 2021). Agentic stereotypes of Black women may thus not map onto traits expected of leaders in a way that enhances their support in ecologically valid contexts. Indeed, Black individuals are often stereotyped negatively on competence (Devine, 1989), an agentic trait characteristic of leaders (Eagly & Karau, 2002a). Some stereotypes of Black women may further limit their attaining leadership roles. Indeed, to the extent that anger is associated with dominance (Knutson, 1996), the "angry Black woman" stereotype contributes to Black women having worse assessments of leadership capability than White women (Motro et al., 2021). These patterns reflect media coverage of real women. For example, media coverage of Stacey Abrams during the Georgia Democratic gubernatorial primary in 2018 contained more negative and agentic stereotypes than coverage of Stacey Evans, a White woman seeking the nomination (Hicks, 2022). A second, and perhaps more likely, possibility is that perceived communality may have a strong positive relation with evaluations of Kamala Harris. Perceived communality may be especially important to combat agentic penalties conferred to Black women

(Rosette et al., 2016). Recent work (Haines & Stroessner, 2019b) outlines how agentic women can be positively perceived, providing a theoretical rationale for this possibility.

Perceived Role Prioritization and Evaluations of Aspiring Female Leaders

Despite women aspiring to political office often being perceived as highly agentic (Koenig et al., 2011), highly agentic perceptions do not always prevent women from the communal perceptions likely key to their success (Okimoto & Brescoll, 2010). Agentic female leaders also perceived as communal often receive an evaluative advantage relative to other leaders (Rosette & Tost, 2010). Such an advantage is consistent with a role prioritization perspective building on role congruity theory. This perspective proposes that whereas perceived low prioritization of traditional gender roles elicits negative evaluations, perceived prioritization of agentic *and* communal behavior may prevent such negativity (Haines & Stroessner, 2019b). If agentic women are perceived as fulfilling their communal roles, they may not be penalized for their agency in other domains. Directly supporting this assertion, unambiguous information about an agentic female leader's communal attributes conveyed through behavior and role information mitigates negativity toward agentic female leaders (Heilman & Okimoto, 2007). Because gender role violations strongly elicit negative attitudes toward female politicians (Schneider et al., 2022), an examination of effects of gendered trait inferences on evaluations of Harris interpreted within a role prioritization framework is especially timely.

Work on actual female leaders indirectly supports this assertion. For example, although Hillary Clinton was perceived as more agentic than communal, only her perceived communality positively related to endorsing her in 2008 (Gervais & Hillard, 2011). The extent of agentic women's perceived communality may thus be especially important for them to be positively evaluated. Because she held office in a male-dominated Senate, we expected people to perceive

Harris as more agentic than communal. Because communality positively relates to stereotypic femininity (Abele, 2003; Eagly & Karau, 2002a; A. E. Martin & Slepian, 2020a) in line with gender prescriptions, we also expected communal inferences to positively relate to evaluations of Harris. Showing such patterns would provide an ecologically valid link to theory positing conditions under which agentic women can "have it all" (Haines & Stroessner, 2019b). It would also expand the social psychological literature on this topic by investigating a real woman of color for the first time.

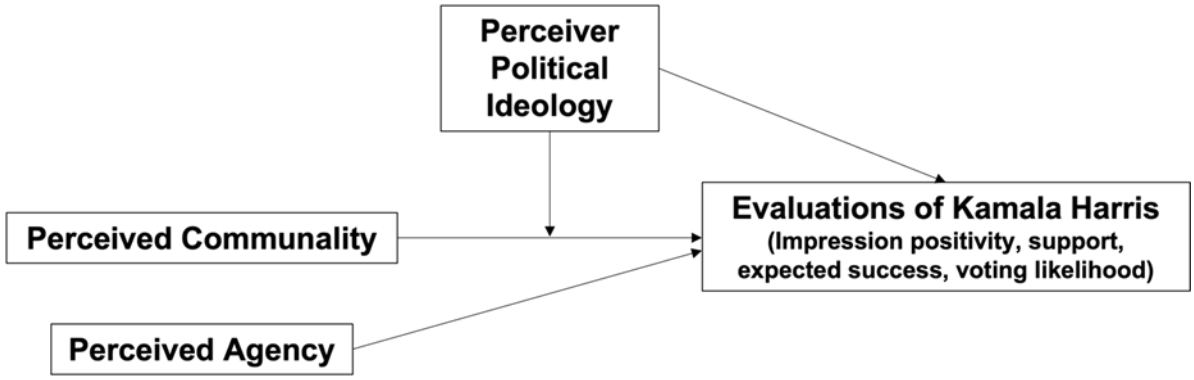
Perceiver Political Ideology Will Qualify the Relation between Gendered Trait Inferences and Evaluations of Aspiring Female Leaders

Past work provides reason to expect that perceived communality will positively relate to evaluations of Kamala Harris. However, these evaluations are likely be affected by the relative political ideology of perceivers because as a Democrat, Harris is perceived as being ideologically liberal. Across the ideological spectrum, people display similar levels of partisan bias, or a tendency to evaluate information more favorably when it supports one's political beliefs (Ditto et al., 2019). This bias extends from how people evaluate information to how they evaluate others. Indeed, liberals and conservatives are similarly intolerant toward others who are ideologically dissimilar from them (Brandt et al., 2014), making negative attributions toward people whose values are inconsistent with their own (Morgan et al., 2010). These findings complement an identity-based (Tajfel, 1970) model of political ideology (Van Bavel & Pereira, 2018) in that people favor candidates sharing their ideology and values over candidates who do not (Iyengar & Krupenkin, 2018; Mallinas et al., 2018; West & Iyengar, 2020).

Social psychological research examining evaluations of actual female politicians has largely examined separable effects of communality and ideology. For example, both perceived

communality and being more ideologically liberal positively related to the likelihood of voting for Hillary Clinton in 2008 even when controlling for sexism (Gervais & Hillard, 2011). Examining an interactive relation between these factors may be critical to characterize communality effects on evaluations of female candidates. Such an examination is warranted given that the salience of partisanship may limit the use of gender stereotypes when evaluating female candidates (Hayes, 2011). An identity-based model of political ideology (Van Bavel & Pereira, 2018) suggests that communality effects on evaluations of Harris may differ by whether perceivers identify as more conservative or more liberal. As a liberal, Harris is more positively evaluated by more liberal than conservative perceivers. One potential outcome of liberal perceivers' more positive evaluations is that Harris may be evaluated positively even if she is perceived as lower in communality. We thus expected a weaker positive relation between communality and evaluations of Harris among more liberal than conservative perceivers. Such a pattern would align with work showing that people show favoritism toward ideologically similar others (Balliet et al., 2018). It would also provide nuance to work on how communality relates to support for agentic female candidates, which often focuses on overall effects without considering ideology beyond using it as a control variable (e.g., Okimoto & Brescoll, 2010). See Figure 1 for a conceptual model of the hypothesized interactive relation of gendered trait inferences and political ideology on evaluations of Kamala Harris.

Figure 1: Conceptual Model of the Hypothesized Effects Between Gendered Trait Inferences and Political Ideology on Evaluations of Kamala Harris in Study 1.



Study 1

Consistent with work on hypothetical candidates (e.g., Okimoto & Brescoll, 2010), we expected Harris to be perceived as more agentic than communal. Consistent with work showing that perceiving female politicians as communal positively relates to their success (e.g., Conroy et al., 2020; Gervais & Hillard, 2011), we expected that perceiving Harris as more communal would relate to favorable evaluations. We expected political ideology to moderate this relation. Political ideology engenders ingroup favoritism (Iyengar & Krupenkin, 2018; Mallinas et al., 2018; West & Iyengar, 2020), driving polarized affective evaluations of political leaders (Rogowski & Sutherland, 2016). That Harris is ideologically liberal could thus elicit a weaker relation between perceived communality and evaluations of her among more liberal than conservative perceivers. Given their shared ideology, more liberal (relative to conservative) perceivers might be expected to evaluate Harris more positively even when they perceive her as lower in communality.

Although this hypothesis is complemented by an identity-based model of political ideology (Van Bavel & Pereira, 2018), people across the ideological spectrum vary by more than how they ideologically identify. For example, conservatism is related to higher sexism (Christopher & Mull, 2006) and more negative attitudes toward social change (Jost et al., 2003, 2008). Ideology could thus moderate a relation between perceived communality and evaluations of Harris because conservatism relates to negative responses to people violating role prescriptions (Silván-Ferrero & Bustillos López, 2007; Wetherell et al., 2013). Thus, a key question is whether any effects of political ideology reflect the favoritism showed toward ideologically similar others (e.g., Brandt et al., 2014; Finkel et al., 2020) or if they function as a proxy for valuing traditional gender prescriptions. To aid in the interpretation of effects involving political ideology, we sought to verify that ideology had a stronger and more consistent effect on evaluations of Harris when accounting for effects of sexism (Glick & Fiske, 1996) and traditional beliefs about social roles (Baber & Tucker, 2006).

Method

Participants

Power analysis (G*Power; Erdfelder et al., 1996) using $f^2 = 0.15$ targeted 103 participants for a significant regression with seven predictors at 80% power and $\alpha = .05$. We oversampled to account for anticipated exclusions and to increase the likelihood of obtaining a wide range of political ideologies. Overall, 175 people from MTurk participated on October 14, 2020. All participants were United States citizens. We recruited participants from MTurk to obtain a larger range of ages and political ideologies than would be possible using an undergraduate participant pool. We did not screen for whether people planned to vote in the 2020 election. All studies were IRB approved.

Three items were used as exclusion criteria. At the beginning of the study, people were asked if they knew who Kamala Harris was (yes/no). Ten were excluded for answering "No." People were asked to "Select 4" when making trait inferences of Harris. All did so. At the end of the study, people indicated whether they answered questions about Kamala Harris or Hillary Clinton. All indicated Harris. These exclusion criteria yielded an analyzed sample of 165 participants (M_{age} = 38.90 years, SD = 12.71; $M_{years\ of\ education}$ = 15.31, SD = 2.15; 79 identifying as female). One-hundred thirty-six participants identified as White, 12 as Black, 11 as Asian, four as multi-racial, and two indicated that their race was unknown to them. Of the 165 participants, 152 also identified as non-Hispanic.

Procedure

Participants read that the study regarded how people evaluate familiar others and that they would make evaluations about Kamala Harris. After indicating whether they knew who Kamala Harris was, participants saw a picture of Kamala Harris standing next to Joe Biden and read commonly repeated information about her: "Kamala Harris attended Howard University and the University of California, Hastings College of the Law. She has been a prosecutor, District Attorney of San Francisco, and Attorney General of California. In 2016, she became the second Black woman and the first South Asian individual elected to the United States Senate. In January 2019, Kamala Harris announced her candidacy for President of the United States, although she dropped out of the race before the end of the year. In August 2020, Joe Biden announced Kamala Harris as his running mate, making her the third female vice-presidential candidate in United States history. Over the next few slides, we will be asking you about your perceptions of Kamala Harris. There are no right or wrong responses."

Trait Inferences and Evaluations of Kamala Harris

Of interest were trait inferences and evaluations of Harris, grouped as listed here. Groupings were presented at random. Within groupings, items were randomly presented. Although we had overall hypotheses regarding communality and political ideology, we selected a variety of different evaluations because they reflect the polling questions often asked of potential voters. These evaluations included perceiver attitudes (impression positivity), expectations (expected success) and intended behavior (support and voting likelihood), each of which are of interest when examining political campaigns (e.g., Bauer, 2017).

Gender Stereotypicality. One item measured Harris's gender stereotypicality ("Relative to other women, how feminine or masculine do you find Kamala Harris?" rated from -3 [*very feminine*] to 3 [*very masculine*]; M = -0.41, SD = 1.28).

Agency and Communality. Participants evaluated Harris on agentic (assertive, ambitious, independent, strong, determined, and persistent) and communal (cooperative, warm, friendly, caring, compassionate, and sympathetic) traits using a scale ranging from 1 (*not at all*) to 7 (*extremely*). The traits were from a validated database (Diehl et al., 2004) and are widely used examples of agentic and communal traits (e.g., Abele, 2003; Bruckmüller & Abele, 2013; Gervais & Hillard, 2011; Trapnell & Paulhus, 2012). Responses to the agentic traits (Cronbach's α = .89) were averaged to create a composite agency inference. Responses to the communal traits (Cronbach's α = .97) were averaged to create a composite communality inference. Within the inferences was one item asking participants to "Select 4." All participants did so.

Impression Positivity. Two items measured impression positivity ("Overall, how likable do you find Kamala Harris to be?" rated from 1 [*not at all*] to 7 [*extremely*]; "Overall, how negative or positive is your impression of Kamala Harris?" rated from 1 [*extremely negative*] to

7 [*extremely positive*]). Responses were highly correlated, $r(163) = .95$, $p < .001$, and were averaged to create a composite impression positivity score.

Support and Expected Success. One item measured support ("Do you support Kamala Harris's candidacy for Vice President?" rated from 1 [*not at all*] to 7 [*completely*]. Two items measured expected success ("How successful do you think Kamala Harris would be as Vice President?" rated from 1 [*not at all*] to 7 [*extremely*]; Relative to other male Vice Presidents, how successful do you think Kamala Harris would be as Vice President?" rated from 1 [*not at all*] to 7 [*extremely*]). Responses to the success-related items were highly correlated, $r(163) = .94$, $p < .001$, and were averaged to create a composite expected success score.

Voting Likelihood. We measured voting likelihood over three items ("Assuming it were possible, what would your likelihood of voting for Kamala Harris be for [Senator/Vice President/President]?" rated from 1 [*very unlikely*] to 7 [*very likely*]). Responses (Cronbach's $\alpha = .96$) were averaged to create a composite voting likelihood score.

Participant Characterization

Participants completed three questionnaires measuring political ideology, sexism, and social role endorsement in a random order after the above-described groupings. The validity of these measures is supported by intercorrelations (Table 1) in expected directions (e.g., hostile sexism is positively correlated with traditional gender role beliefs). Reliability (measured via Cronbach's α) indicates the reliability of responses within this participant sample.

Table 1: Intercorrelations Between Variables Measured in Study 1

Variable	*M* (*SD*)	1	2	3	4	5	6	7	8	9	10
1. Impression positivity	4.46 (2.08)										
2. Support	4.70 (2.36)	.94** [.92, .95]									
3. Expected success	4.88 (1.93)	.89** [.86, .92]	.90** [.86, .92]								
4. Voting likelihood	4.34 (2.31)	.92** [.90, .94]	.94** [.92, .96]	.87** [.83, .91]							
5. Communal	4.40 (1.72)	.91** [.88, .93]	.84** [.79, .88]	.84** [.79, .88]	.84** [.79, .88]						
6. Agentic	5.94 (1.01)	.40** [.26, .52]	.45** [.32, .56]	.50** [.38, .61]	.45** [.32, .57]	.47** [.34, .58]					
7. Ideology	5.86 (2.30)	.65** [.55, .73]	.70** [.61, .77]	.71** [.62, .78]	.67** [.58, .75]	.63** [.53, .71]	.33** [.19, .46]				
8. Hostile sexism	1.61 (1.26)	-.24** [-.38, -.09]	-.29** [-.43, -.15]	-.36** [-.49, -.22]	-.33** [-.46, -.19]	-.25** [-.39, -.10]	-.29** [-.42, -.14]	-.44** [-.56, -.31]			
9. Benevolent sexism	1.85 (1.18)	.05 [-.10, .20]	-.03 [-.18, .13]	-.08 [-.23, .07]	-.04 [-.19, .11]	.10 [-.06, .25]	-.05 [-.20, .10]	-.22** [-.36, -.06]	.46** [.33, .57]		
10. Non-Gender transcendent beliefs	18.19 (18.18)	-.11 [-.26, .04]	-.17* [-.32, -.02]	-.19* [-.34, -.04]	-.22** [-.36, -.07]	-.12 [-.27, .03]	-.40** [-.52, -.26]	-.25** [-.39, -.10]	.39** [.25, .51]	.36** [.22, .48]	
11. Gender linked beliefs	32.63 (24.71)	-.14 [-.29, .01]	-.20* [-.34, -.05]	-.25** [-.38, -.10]	-.21** [-.35, -.06]	-.11 [-.26, .05]	-.16* [-.31, -.01]	-.36** [-.48, -.22]	.67** [.57, .74]	.62** [.51, .70]	.46** [.33, .57]

Note. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * $p < .05$. ** $p < .01$.

Political Ideology. Participants indicated political ideology over four items (overall, economic issues, social issues, and foreign policy issues) on a scale ranging from 1 (*extremely conservative*) to 9 (*extremely liberal*). Responses (Cronbach's α = .97) were averaged to create a composite political ideology score. Note that although relative ideology does not exactly match with the American political parties (i.e., Democrats and Republicans), these correlated concepts (Abramowitz & Saunders, 2008) can determine partisanship effects on person perception (e.g., Wilson & Rule, 2014).

Sexism. Participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). The ASI measures hostile (e.g., "Women are too easily offended") and benevolent (e.g., "Women should be cherished and protected by men") sexism on a scale ranging from 0 (*disagree strongly*) to 5 (*agree strongly*). Higher scores indicate more sexism. Items measuring hostile (Cronbach's α = .93) and benevolent (Cronbach's α = .87) sexism were averaged to create hostile and benevolent sexism scores.

Social Role Beliefs. Participants completed the Social Roles Questionnaire (SRQ; Baber & Tucker, 2006) to measure gender-transcendent (e.g., "Tasks around the house should not be assigned by sex") and gender-linked (e.g., "For many important jobs, it is better to choose men instead of women") beliefs. Participants indicated how much they agreed with each statement from 0% to 100% in 10% increments. We reverse-coded gender-transcendent items so they could be interpreted like the gender-linked items (i.e., higher values measuring non-transcendent gender beliefs). Higher scores reflect more traditional social role beliefs. Items measuring non-transcendent gender beliefs (Cronbach's α = .80) and items measuring gender-linked beliefs (Cronbach's α = .91) were averaged to create two social role belief scores.

Results

Kamala Harris is Perceived as Being More Agentic than Communal

A paired samples t-test showed that people perceived Harris as more agentic (M = 5.94, SD = 1.01) than communal (M = 4.40, SD = 1.72), $t(164) = 12.99$, $p < .001$, $d = 1.01$. To show gendered trait inferences, we regressed gender stereotypicality ratings on composite communality and agency (each standardized around the overall mean). We included standardized composite ideology scores because ideology might be expected to broadly affect perceptions. The model was significant, $F(3, 161) = 7.65$, $p < .001$, $R^2 = .12$. Perceptions of greater communality related to perceiving Kamala Harris as more feminine, $b = -0.50$, $SE = 0.13$, $t = 3.92$, $p < .001$. No significant relations emerged with perceived agency, $b = 0.11$, $SE = 0.11$, $t = 1.01$, $p = .31$, or ideology, $b = 0.03$, $SE = 0.12$, $t = 0.22$, $p = .83$. A model including Participant Gender (male = -1 and female = 1) did not explain more variance, $F(1,160) = 1.10$, $p = .29$.

Interactive Effects of Perceived Communality and Political Ideology on Evaluations of Kamala Harris

We conducted a multivariate multiple regression by entering standardized composite communality inferences, composite agency inferences, composite ideology scores, and their interactions simultaneously into a model. This model yielded Communality (Pillai's trace = 0.73, $F(4, 154) = 102.58$, $p < .001$), Agency (Pillai's trace = 0.20, $F(4, 154) = 9.37$, $p < .001$), and Ideology (Pillai's trace = 0.28, $F(4, 154) = 14.73$, $p < .001$) effects. The expected Communality × Ideology interaction emerged, Pillai's trace = 0.22, $F(4, 154) = 11.04$, $p < .001$. An unexpected Agency × Ideology interaction emerged, Pillai's trace = 0.10, $F(4, 154) = 4.29$, $p = .003$. There was no Communality × Agency interaction, Pillai's trace = 0.05, $F(4, 154) = 2.13$, $p = .08$, and no three-way interaction, Pillai's trace = 0.03, $F(4, 154) = 1.29$, $p = .28$.

To characterize these effects, we regressed the same predictors on each evaluation. We examined interactions involving ideology using values one standard deviation above (more liberal) and below (more conservative) the mean composite political ideology score. Models including Participant Gender (male = -1 and female = 1) and its interactions did not explain significantly more variance in the models than the more parsimonious models, Fs < 1.64, ps > .12. Participant Gender was thus not explored further. We made an a priori decision not to aggregate across the four evaluations because they reflect different kinds of polling questions asked of potential voters. See the Supplemental Material (A) for analyses aggregating across them.

Impression Positivity

The model was significant, $F(7, 157) = 122.00$, $p < .001$, $R^2 = .84$ (Table 2a). More communality, $b = 1.77$, $SE = 0.10$, $t = 18.19$, $p < .001$, and having a more liberal ideology, $b = 0.23$, $SE = 0.10$, $t = 2.43$, $p = .02$, was related to more impression positivity. The expected interaction between communality and political ideology did not emerge, $b = -0.09$, $SE = 0.08$, $t = 1.12$, $p = .27$.

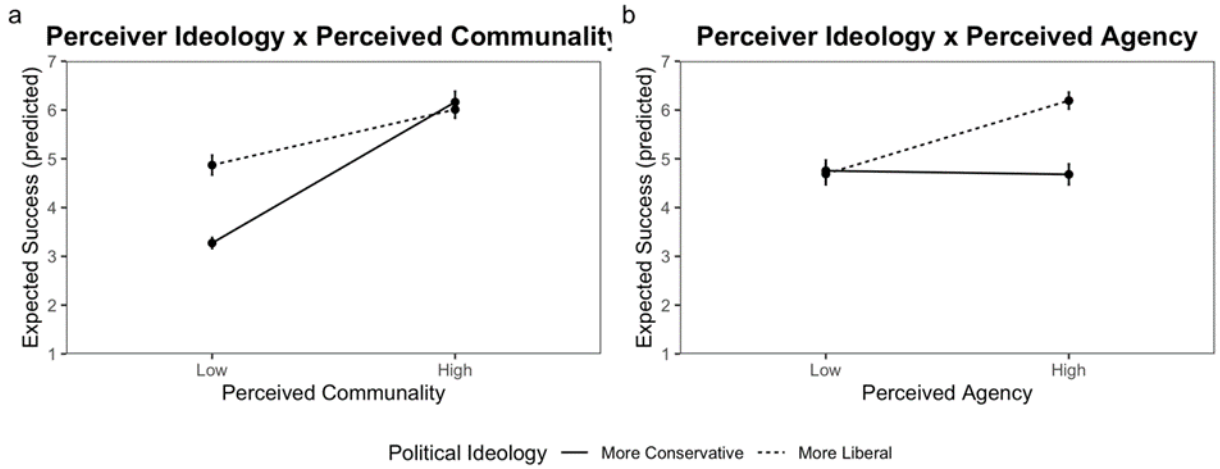
Support

The model was significant, $F(7, 157) = 74.60$, $p < .001$, $R^2 = .77$ (Table 2b). More communality, $b = 1.45$, $SE = 0.14$, $t = 10.75$, $p < .001$, and having a more liberal ideology, $b = 0.60$, $SE = 0.13$, $t = 4.55$, $p < .001$, was related to more support. The expected interaction between communality and political ideology did not emerge, $b = -0.19$, $SE = 0.11$, $t = 1.69$, $p = .09$.

Expected Success

The model was significant, $F(7, 157) = 97.24$, $p < .001$, $R^2 = .81$ (Table 2c). More communality, $b = 1.01$, $SE = 0.10$, $t = 10.14$, $p < .001$, and having a more liberal ideology, $b = 0.36$, $SE = 0.11$, $t = 3.73$, $p < .001$, was related to more expected success. The expected Communality × Ideology interaction emerged, $b = -0.43$, $SE = 0.08$, $t = 5.30$, $p < .001$ (Figure 2a). Perceived communality had a stronger positive relationship with expected success for more conservative, $b = 1.45$, $SE = 0.11$, $t = 13.47$, $p < .001$, than liberal, $b = 0.57$, $SE = 0.15$, $t = 3.85$, $p < .001$, participants.

Figure 2: Interactive Effects on Expected Success.



Note. In Study 1, perceived communality had a stronger relation with expected success among more conservative than more liberal participants (a). Perceived agency had a stronger relation with expected success among more liberal than conservative participants (b).

Agency related to expected success, $b = 0.36$, $SE = 0.10$, $t = 3.40$, $p < .001$. An unexpected Agency × Ideology interaction also emerged, $b = 0.17$, $SE = 0.05$, $t = 3.24$, $p = .001$ (Figure 2b). Perceived agency was positively related to expected success for more liberal, $b =$

0.75, *SE* = 0.16, *t* = 4.80, *p* < .001, but not conservative participants, *b* = -0.04, *SE* = 0.17, *t* = 0.23, *p* = .82.

Table 2: Regression Summaries for Evaluations of Kamala Harris by Gendered Trait Inferences and Political Ideology in Study 1 (N = 165).

	Impression Positivity			**Support**			**Expected Success**			**Voting Likelihood**		
	b (SE)	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*
Communality	1.77 (0.10)	18.19	< .001	1.45 (0.14)	10.75	< .001	1.01 (0.08)	10.14	< .001	1.53 (0.14)	11.13	< .001
Agency	-0.13 (0.10)	1.22	.23	0.29 (0.14)	2.03	0.04	0.36 (0.11)	3.40	< .001	0.26 (0.14)	1.77	.08
Ideology	0.23 (0.10)	2.43	.02	0.60 (0.13)	4.55	< .001	0.36 (0.10)	3.73	< .001	0.60 (0.13)	4.44	< .001
Communality x Agency	-0.13 (0.08)	1.64	.10	-0.04 (0.11)	0.33	.74	-0.11 (0.08)	1.45	.14	-0.02 (0.11)	0.21	.83
Communality x Ideology	-0.09 (0.08)	1.12	.27	-0.19 (0.11)	1.69	.09	-0.44 (0.08)	5.30	< .001	0.04 (0.11)	0.33	.74
Agency x Ideology	0.07 (0.12)	0.62	.54	0.06 (0.17)	0.34	.73	0.39 (0.12)	3.24	.001	-0.02 (0.17)	0.12	.91
Communality x Agency x Ideology	-0.03 (0.07)	0.44	.66	-0.11 (0.10)	1.08	.28	0.05 (0.07)	0.75	.46	-0.13 (0.10)	1.27	.20

Voting Likelihood

The model was significant, $F(7, 157) = 66.52$, $p < 0.001$, $R^2 = 0.75$ (Table 2d). More communality, $b = 1.53$, $SE = 0.11$, $t = 11.13$, $p < .001$, and having a more liberal ideology, $b = 0.60$, $SE = 0.13$, $t = 4.44$, $p < .001$, was related to a higher voting likelihood. The expected interaction between communality and political ideology did not emerge, $b = 0.04$, $SE = 0.11$, $t = 0.33$, $p = .74$.

Interpreting Political Ideology Effects

Two interpretations for ideology effects on evaluations seemed plausible. First, given the identity-based theoretical account of political ideology (Van Bavel & Pereira, 2018), ideology might positively relate to evaluations because more liberal perceivers shared aspects of identity and values with Harris. Second, because ideological conservatism relates to negative responses to people violating role prescriptions (Silván-Ferrero & Bustillos López, 2007; Wetherell et al., 2013), conservatives might have more negatively evaluated Harris because she violated the female role prescription through her high agency. To aid in the interpretation of effects involving political ideology, we next regressed evaluations of Harris (impression positivity, support, expected success, and voting likelihood) on political ideology, hostile sexism, benevolent sexism, non-transcendent gender beliefs, and gender-linked beliefs (all standardized). More liberal ideology had significant positive relations with evaluations across models. No other significant and consistent relations emerged.

These findings do not suggest that political ideology is the only predictor of evaluations of Harris. However, they suggest that ideology provides strongly consistent relations beyond relations with variables more face valid in measuring sexism and gender role beliefs. We thus interpreted our findings regarding political ideology from an identity-based perspective (Van

Bavel & Pereira, 2018). If ideology was a proxy for valuing traditional gender prescriptions, we would expect the other variables to also have consistently significant positive relations with evaluations or for relations to be more comparable to the strength of the ideology effect.

Discussion

People perceived Kamala Harris as being more agentic than communal, paralleling work using hypothetical (e.g., Okimoto & Brescoll, 2010) and actual (e.g., Gervais & Hillard, 2011) candidates. Like past work (Abele et al., 2008; Abele & Wojciszke, 2007; Martin & Slepian, 2020), perceiving Harris as more communal related to perceiving her as more feminine relative to other women. Perceptions of agency did not significantly relate to her evaluated gender stereotypicality. This finding might seem surprising because agency is associated with masculinity (Abele, 2003; Bem, 1974). Speculatively, evaluating Harris relative to other women may have increased the salience of communality as a predictor of gender stereotypicality. Like related work (Gervais & Hillard, 2011), we measured gender stereotypicality using one item. Because recent work suggests that masculinity and femininity are distinct dimensions in person perception (Hester et al., 2021), future work may consider multiple items when examining evaluations of women to link these literatures.

Perceived communality positively related to evaluations of Harris even when accounting for perceived agency and political ideology. This finding reflects that communal perceptions are especially important for women to be favorably evaluated (e.g., Eagly & Karau, 2002). This pattern may link to work showing that women who are perceived as agentic *and* communal often receive evaluative advantages relative to leaders with other combinations of these traits (Rosette & Tost, 2010), supporting a role prioritization perspective (Haines & Stroessner, 2019b).

The multivariate analysis showed the expected interactive relation between perceiver political ideology and perceived communality on evaluations of Harris. Univariate analyses, however, only showed an interactive effect for Harris's expected success. Expected success may differ from the other evaluations in that it speaks to Harris in relation to a leadership role rather than attitudes and behaviors toward Harris herself. The positive relation between perceived communality and expected success was attenuated among more liberal than conservative perceivers. This pattern was due to more liberal than conservative perceivers expecting more success when Harris was perceived as less but not more communal. From an identity-based perspective, this pattern is consistent with ingroup favoritism (Tajfel, 1970), the current polarized political climate (Finkel et al., 2020) and work showing that ideology polarizes impressions of partisans (Mallinas et al., 2018; Rogowski & Sutherland, 2016). Sharing an ideology with female candidates (relative to not) may buffer success expectations when women are perceived as less communal.

Unexpectedly, perceived agency and ideology interacted to affect expected success. Here, the positive relation between perceived agency and expected success was stronger for more liberal than conservative perceivers. This pattern was due to more liberal relative to conservative perceivers expecting more success when Harris was perceived as more, but not less, agentic. Because agentic traits characterize leaders (Eagly & Karau, 2002a), one possibility is that being perceived as more agentic may yield evaluations that Harris is more likely to be successful overall. Speculatively, sharing a female candidate's ideology may augment expectations of her success given any positive evaluations. Future work may directly address this possibility by manipulating candidate ideology.

Although the multivariate analysis suggested an interaction between perceived communality and ideology, this pattern did not reach the statistical significance across evaluations. One possibility is that the description of Harris reflected her prioritization of agency by describing her education and professional accomplishments. The role prioritization model (Haines & Stroessner, 2019b), however, posits that agentic women perceived as prioritizing communality are less penalized for their agency than women who do not prioritize it (Heilman & Okimoto, 2007; Rosette & Tost, 2010a). Liberals could have a weaker communality effect on evaluations of Harris than conservatives when they are provided information describing her communality through role information and behavior. By contrast, receiving information describing Harris's agency may make perceiving communality more broadly important for perceivers to positively evaluate her. Study 2 addresses this possibility by having people read descriptions conveying information about Harris's agency or communality before evaluating her.

Study 2

Although aspiring female leaders are often penalized for perceived low communality (Okimoto & Brescoll, 2010; Schneider et al., 2022), women who have achieved top-level positions may receive evaluative advantages if they are perceived as highly agentic *and* as prioritizing communality (Heilman & Okimoto, 2007; Rosette & Tost, 2010). One possibility is that if people receive agentic information about a high-achieving agentic woman, perceived communality might be an especially important contributor to positive evaluations of her. Recent work indirectly supports this possibility.

In one study (Cassidy & Liebenow, 2021a), conveyed agentic or communal information about Harris was manipulated between-participants. People then selected which face out of an array best resembled her. These faces were manipulated such that some were perceived as more

communal than agentic. A positive relation between selecting a communal face and positive evaluations of Harris only emerged when her agency was described. This relation may reflect that perceived communality is broadly important for agentic women to be positively evaluated when given information about her conveys agency. That the description of Harris in Study 1 conveyed her agency through educational and professional accomplishments and roles raises the possibility that communal evaluations were especially important to elicit positive evaluations of her in this context. If true, an interaction between ideology and perceived communality could be blunted based on whether a given description of Harris conveys agentic information.

To test for this possibility, we re-ran Study 1 and modified the task in two ways. First, we manipulated the description of Harris between-participants. The description conveyed details about Harris's roles and behaviors speaking to her agency (e.g., education and professional information) or communality (e.g., relationships and family life). Second, people made agentic and communal trait inferences immediately after reading the description. This modification directly linked trait inferences to the descriptions and prevented inferences from potentially being affected by other evaluations. Although we expected Harris to be perceived as more agentic than communal overall, we expected a larger difference among people who read the description conveying more agentic versus more communal information.

If a description conveyed information regarding Harris's more agentic roles and behaviors, we reasoned that communal inferences should be especially important for Harris not to receive an agentic penalty (i.e., more negative evaluations). We thus expected that the description of Harris would qualify how political ideology and perceived communality interact to affect evaluations. If information about Harris's agentic roles and behaviors makes perceived communality especially salient, similarly positive communality effects on evaluations across

more liberal and conservative perceivers should emerge. By contrast, we expected more liberal perceivers to show a weaker positive communality effect on evaluations than more conservative perceivers when a description focused on Harris's more communal roles and behaviors. Evidence of communal roles and behavior may buffer negative effects of perceiving low communality to the extent that people share ideology with a candidate.

Method

Participants

We recruited 275 MTurk participants who did not complete Study 1 on October 16, 2020. Exclusion criteria mirrored Study 1. Nine participants were excluded for indicating that they did not know who Kamala Harris was. Two were excluded for not passing the "Select 4" attention check. One was excluded for indicating they evaluated Hillary Clinton. These exclusions yielded an analyzed sample of 263 participants (M_{age} = 40.60 years, SD = 12.136; $M_{years\ of\ education}$ = 15.52, SD = 2.15; 124 identifying as female). Two-hundred eighteen participants identified as White, 20 as Black, 14 as Asian, two as Native American/Alaska Native, six as multi-racial, and two reported having an unknown race. One did not report race. Of these participants, 237 identified as non-Hispanic. All participants were United States citizens. We did not screen for whether participants planned to vote in the 2020 election.

Procedure

Study 2 replicated Study 1 with the following changes. Participants were randomly assigned to read a description of Harris conveying information about her agentic (n = 127) or communal (n = 136) roles and behaviors for at least 30 seconds. Participants reading the agentic (M = 79.87s, SD = 99.12) and communal (M = 87.94s, SD = 202.38) description did not differ in

how much time they spent on the description screen, $t(261) = 0.41$, $p = .67$, $d = 0.05$. Male and female participants were similarly represented across descriptions, $\chi^2(1) = 0.16$, $p = .69$.

For the description conveying agentic information, participants saw a picture of Kamala Harris giving a speech at a podium and read, “Kamala Harris attended Howard University and the University of California, Hastings College of the Law. She has been a prosecutor, District Attorney of San Francisco, and Attorney General of California. As California's attorney general, Harris prioritized the prosecution of transnational gangs and weapons, drugs, and human traffickers. She extracted $25 billion from Wall Street to compensate California homeowners following the financial crisis. Additionally, she won a $1.1 billion settlement against a for-profit-college chain over allegations of predatory and illegal practices. In 2016, she became the second Black woman and the first South Asian individual elected to the United States Senate. Since then, she has become a prominent voice in Congress. In January 2019, Kamala Harris announced her candidacy for President of the United States, although she dropped out of the race before the end of the year. In August 2020, Joe Biden announced Kamala Harris as his running mate, making her the third female vice-presidential candidate in United States history.”

For the description conveying communal information, participants saw a picture of Kamala Harris smiling at her wedding with family members and read, “Kamala Harris was born and raised in Oakland, California to her parents Shyamala and Donald. After Kamala's parents divorced when she was seven years old, her mother primarily raised her and her sister. Kamala is very close to her younger sister, Maya Harris, who was the campaign chairwoman for her presidential bid. Kamala's niece, Meena, is also very close to her and wrote a children's book about Kamala and Maya's childhood. Kamala has been married to Douglas Emhoff, a lawyer, since 2014. Kamala is a stepmother to Cole and Ella, with whom she is close. Both endearingly

refer to Kamala as "Momala." Kamala has a tradition of weekly Sunday dinners with her family, where she makes dinner for family and friends. In August 2020, Joe Biden announced Kamala as his running mate, making her the third female vice-presidential candidate in United States history."

Note that these descriptions differed in conveyed agency and communality and in the context in which information is presented. Whereas the agentic description focuses on Harris's roles and behaviors in a public setting, the communal description does so in the context of her private life. These differences reflect media portrayal of how female political candidates convey their communality (Carlin & Winfrey, 2009a) and speak to work showing that communal traits shown in agentic women's personal lives affect perceptions of their job performance (Bligh et al., 2012a; Diekman & Schneider, 2010; Rosette & Tost, 2010a). Future work may consider how communality conveyed through information about agentic women's public lives affects evaluations of them. In this context, their agentic leadership positions would still be highlighted.

Participants evaluated agentic and communal traits after reading the description. Agentic inferences (Cronbach's $\alpha = .88$) were averaged to create a composite agency inference. Communal inferences (Cronbach's $\alpha = .97$) were averaged to create a composite communality inference.

Evaluations of Kamala Harris

Participants next completed the other groupings from Study 1 in a random order. Responses to the impression positivity items, $r(261) = .93$, $p < .001$, were averaged to create a composite score. Responses to the success-related items, $r(261) = .94$, $p < .001$, were averaged to create a composite score. Responses to voting likelihood items (Cronbach's $\alpha = .96$) were averaged to create a composite score.

Participant Characterization

Participants then completed the questionnaires from Study 1 in a random order. Responses to political ideology items (Cronbach's $\alpha = .96$) were averaged to create a composite score. For the ASI, items measuring hostile (Cronbach's $\alpha = .94$) and benevolent (Cronbach's $\alpha = .89$) sexism were averaged to make composite scores. For the SRQ, items measuring non-transcendent gender beliefs (Cronbach's $\alpha = .78$) and gender egalitarian beliefs (Cronbach's $\alpha = .90$) were averaged to make composite scores. The validity of these measures is again supported by intercorrelations (Table 3) in expected directions. No differences emerged by description, *t*s < 1.14, *p*s > .26.

Table 3: Intercorrelations Between Variables Measured in Study 2

Variable	*M* (*SD*)	1	2	3	4	5	6	7	8	9	10
1. Impression positivity	4.74 (1.97)										
2. Support	4.75 (2.37)	.90** [.87, .92]									
3. Expected success	5.00 (1.89)	.87** [.83, .89]	.87** [.84, .90]								
4. Voting likelihood	4.44 (2.34)	.88** [.85, .90]	.95** [.93, .96]	.86** [.83, .89]							
5. Communal	4.93 (1.53)	.82** [.78, .86]	.76** [.71, .81]	.76** [.70, .80]	.75** [.69, .79]						
6. Agentic	5.85 (1.00)	.49** [.39, .58]	.52** [.42, .60]	.58** [.49, .65]	.51** [.42, .60]	.52** [.43, .60]					
7. Ideology	5.82 (2.28)	.65** [.57, .71]	.68** [.61, .74]	.64** [.56, .70]	.70** [.63, .76]	.54** [.45, .62]	.39** [.29, .49]				
8. Hostile sexism	1.68 (1.26)	-.41** [-.50, -.30]	-.47** [-.56, -.37]	-.51** [-.59, -.41]	-.49** [-.58, -.40]	-.35** [-.45, -.23]	-.35** [-.45, -.24]	-.38** [-.48, -.27]			
9. Benevolent sexism	2.06 (1.17)	-.12* [-.24, -.00]	-.16* [-.27, -.04]	-.14* [-.26, -.02]	-.17** [-.29, -.05]	-.02 [-.14, .11]	-.06 [-.18, .06]	-.32** [-.42, -.20]	.36** [.25, .46]		
10. Non-Gender transcendent beliefs	19.66 (18.64)	-.29** [-.40, -.18]	-.36** [-.46, -.25]	-.40** [-.50, -.30]	-.38** [-.48, -.27]	-.32** [-.42, -.20]	-.37** [-.47, -.26]	-.31** [-.41, -.19]	.47** [.38, .56]	.20** [.08, .31]	
11. Gender linked beliefs	37.35 (24.36)	-.21** [-.33, -.10]	-.28** [-.38, -.16]	-.32** [-.42, -.21]	-.30** [-.40, -.18]	-.12 [-.23, .00]	-.14* [-.26, -.02]	-.35** [-.45, -.24]	.65** [.57, .71]	.56** [.47, .64]	.45** [.35, .54]

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

Results

Agentic and Communal Perceptions of Kamala Harris Vary by Description Information

To test whether a description conveying information about Harris's agency or communality affected trait inferences, we subjected composite agency and communality inferences to a 2 (Description: agentic, communal) × 2 (Trait: agency, communality) mixed model ANOVA. Replicating Study 1, a Trait effect showed that Harris was perceived as more agentic (M = 5.86, SD = 1.00) than communal (M = 4.93, SD = 1.53), $F(1, 261) = 153.63$, $p < .001$, $\eta_p^2 = .37$. A Description × Trait interaction emerged, $F(1, 261) = 41.42$, $p < .001$, $\eta_p^2 = .14$. With the agentic description, people perceived Harris as more agentic (M = 6.02, SD = 1.01) than communal (M = 4.59, SD = 1.61), $t(126) = 12.10$, $p < .001$, $d = 1.07$. With the communal description, people perceived Harris as more agentic (M = 5.70, SD = 0.97) than communal (M = 5.25, SD = 1.39) to a lesser extent, $t(135) = 4.68$, $p < .001$, $d = 0.40$. That the communal description did not override perceptions that Harris is more agentic speaks to the strength of agentic perceptions of her. The smaller difference given the communal description was due to Harris being perceived as more agentic given the agentic relative to the communal description, $t(261) = 2.60$, $p < .001$, $d = 0.32$, and as more communal given the communal relative to the agentic description, $t(261) = 3.57$, $p < .001$, $d = 0.44$. No Description effect emerged, $F(1, 261) = 1.54$, $p = .22$, $\eta_p^2 < .01$, likely because of the strong overall perception that Harris is highly agentic.

We again regressed evaluated gender stereotypicality on standardized composite communality and agency inferences as well as standardized composite ideology scores. The model was significant, $F(3, 259) = 13.54$, $p < .001$, $R^2 = .14$. More perceived communality related to perceiving Harris as more feminine, b = -0.56, SE = 0.10, t = -5.38, $p < .001$. Having a

more liberal ideology related to perceiving Harris as more masculine, $b = 0.24$, $SE = 0.10$, $t = 2.44$, $p = .02$. The agency effect was non-significant, $b = -0.76$, $SE = 0.09$, $t = -0.80$, $p = .42$. Including Participant Gender (male = -1 and female = 1) did not explain more variance than the first model, $F(1, 258) = 0.12$, $p = .73$. Including Description (agentic = -1 and communal = 1) did not explain more variance than the first model, $F(1, 258) = 0.09$, $p = .76$.

Interactive Effects of Perceived Communality and Political Ideology on Evaluations of Kamala Harris

We conducted a multivariate multiple regression entering standardized composite communality inferences, composite agency inferences, composite ideology scores, Description (agentic = -1, communal = 1) and their interactions simultaneously into a model (Table 4). Like Study 1, this model yielded Communality (Pillai's trace = 0.50, $F(4, 244) = 61.34$, $p < .001$), Agency (Pillai's trace = 0.09, $F(4, 244) = 5.98$, $p < .001$), and Ideology (Pillai's trace = 0.28, $F(4, 244) = 23.89$, $p < .001$) effects. Like Study 1, this model yielded a Communality × Ideology interaction, Pillai's trace = 0.05, $F(4, 244) = 2.91$, $p = .02$. This model uniquely yielded interactions between Description, Communality, and Ideology (Pillai's trace = 0.04, $F(4, 244) = 2.62$, $p = .04$) and between Communality, Agency, and Ideology (Pillai's trace = 0.09, $F(4, 244) = 6.18$, $p < .001$).

Regressions on each evaluation characterized these relations. We examined interactions involving composite political ideology using values one standard deviation above (more liberal) and below (more conservative) the mean composite political ideology score. Exploratory models including Participant Gender (male = -1 and female = 1) and its interactions did not explain more variance than the more parsimonious models, Fs < 1.13, ps > .32.

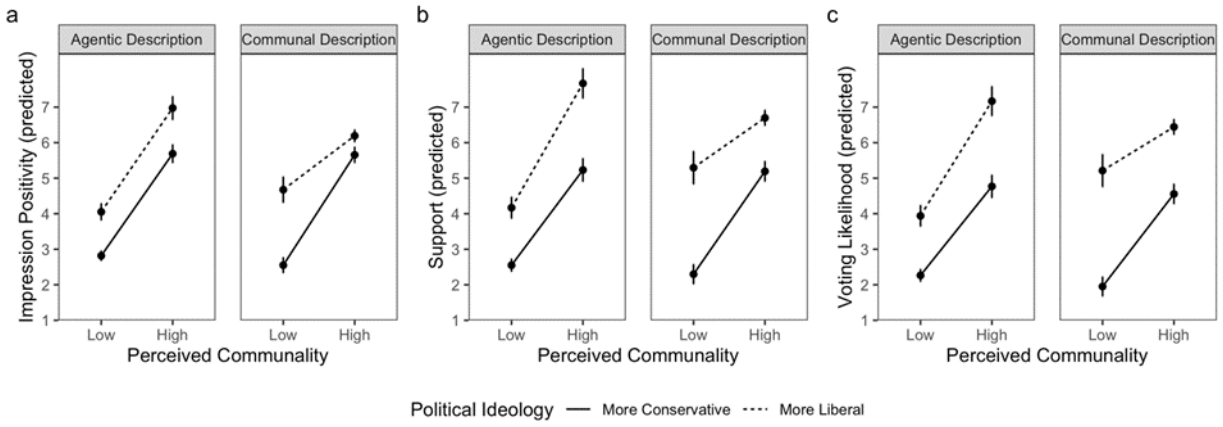
Table 4: Multivariate Regression Summary for Evaluations of Kamala Harris by Description, Gendered Trait Inferences, and Political Ideology in Study 2 (N = 263).

	Pillai's trace	*F*	*p*
Communality	0.50	61.34	< .001
Agency	0.09	5.98	< .001
Ideology	0.28	23.89	< .001
Description	0.02	1.03	.39
Communality x Agency	0.04	2.26	.06
Communality x Ideology	0.05	2.91	.03
Agency x Ideology	0.01	.75	.56
Communality x Description	0.006	.38	.82
Agency x Description	0.008	.50	.74
Ideology x Description	0.02	.97	.42
Communality x Agency x Ideology	0.09	6.18	< .001
Communality x Agency x Description	.02	1.17	.33
Communality x Ideology x Description	.04	2.62	.04
Agency x Ideology x Description	.01	.65	.63
Communality x Agency x Ideology x Description	.001	.04	.99

Impression Positivity

The model was significant, $F(15, 247) = 51.65$, $p < .001$, $R^2 = .76$ (Table 5a). Positive Communality, $b = 1.30$, $SE = 0.10$, $t = 13.42$, $p < .001$, and Ideology, $b = 0.65$, $SE = 0.09$, $t = 7.36$, $p < .001$, effects emerged. The expected Communality × Ideology × Description interaction emerged (Figure 3a), $b = -0.20$, $SE = 0.08$, $t = 2.45$, $p = .02$. To characterize this interaction, we compared communality effects for more liberal and more conservative participants reading each description. With the agentic description, communality effects similarly emerged among more conservative, $b = 1.44$, $SE = 0.12$, $t = 11.71$, $p < .001$, and liberal, $b = 1.46$, $SE = 0.21$, $t = 6.92$, $p < .001$, participants, $b = -0.02$, $SE = 0.21$, $t = 0.12$, $p = .91$. With the communal description, communality effects were stronger for more conservative, $b = 1.55$, $SE = 0.16$, $t = 9.63$, $p < .001$, than liberal, $b = 0.76$, $SE = 0.21$, $t = 3.54$, $p = .001$, participants, $b = 0.79$, $SE = 0.26$, $t = 3.10$, $p = .002$.

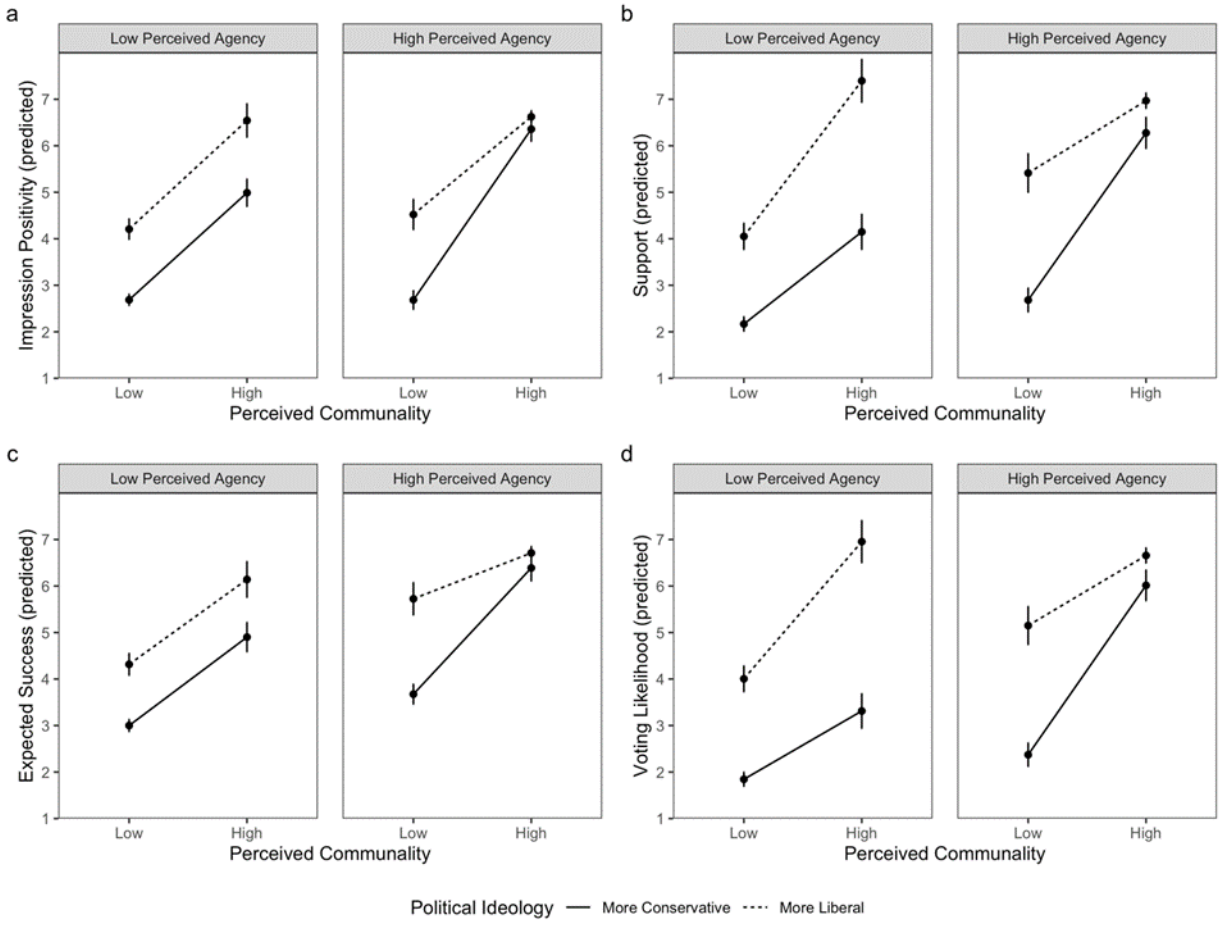
Figure 3: Description x Perceiver x Perceived Communality Interactions



Note: In Study 2, stronger communality effects on impression positivity (a), support (b), and voting likelihood (d) emerged among more conservative than more liberal participants when a description conveyed role and behavior information about Harris reflecting her agency, but not her communality.

An Agency × Communion × Ideology interaction also emerged (Figure 4a), b = -0.20, SE = 0.07, t = 2.71, p = .007. To characterize this interaction, we compared communality effects for more liberal and more conservative participants one standard deviation below (less agentic) and above (more agentic) the mean agency inference. When Harris was perceived as less agentic, communality effects similarly emerged among more conservative, b = 1.15, SE = 0.13, t = 9.03, p < .001, and liberal, b = 1.17, SE = 0.20, t = 5.82, p < .001, participants, b = -0.02, SE = 0.22, t = -0.08, p = .95. When more agentic, communality effects were stronger for more conservative, b = 1.84, SE = 0.16, t = 11.88, p < .001, than liberal, b = 1.05, SE = 0.18, t = 6.01, p < .001, participants, b = 0.79, SE = 0.22, t = 3.525, p = .001.

Figure 4: Perceiver Ideology x Perceived Communality x Perceived Agency Interactions.



Note: In Study 2, stronger communality effects emerged on impression positivity (a), support (b), expected success (c), and voting likelihood (d) for more liberal than for more conservative participants only when Harris was perceived as highly agentic.

Support

The model was significant, $F(15, 247) = 43.54$, $p < .001$, $R^2 = .73$ (Table 5b). Positive Communality, $b = 1.31$, $SE = 0.12$, $t = 10.53$, $p < .001$, Agency, $b = 0.45$, $SE = 0.13$, $t = 3.41$, $p < .001$, and Ideology, $b = 1.07$, $SE = 0.11$, $t = 9.49$, $p < .001$, effects emerged.

The expected Communality × Ideology × Description interaction emerged (Figure 3b), $b = -0.29$, $SE = 0.11$, $t = 2.69$, $p = .008$. With the agentic description, communality effects

similarly emerged among more conservative, b = 1.34, SE = 0.16, t = 8.52, p < .001, and liberal, b = 1.75, SE = 0.27, t = 6.47, p < .001, participants, b = -0.41, SE = 0.27, t = 1.50, p = .13. With the communal description, communality effects were stronger for more conservative, b = 1.45, SE = 0.21, t = 7.01, p < .001, than liberal, b = 0.70, SE = 0.27, t = 2.56, p = .01, participants, b = 0.75, SE = 0.33, t = 2.25, p = .03.

The Agency × Communion × Ideology interaction emerged, (Figure 4b), b = -0.43, SE = 0.09, t = 4.48, p < .001. When Harris was perceived as less agentic, communality effects were stronger for more liberal, b = 1.67, SE = 0.26, t = 6.50, p < .001, than conservative, b = 0.99, SE = 0.16, t = 6.06, p < .001, participants, b = -0.68, SE = 0.29, t = -2.38, p = .02. When more agentic, communality effects were stronger for more conservative, b = 1.80, SE = 0.20, t = 9.07, p < .001, than liberal, b = 0.78, SE = 0.22, t = 3.47, p = .001, participants, b = 1.02, SE = 0.29, t = 3.57, p = .001.

Expected Success

The model was significant, $F(15, 247) = 38.48$, p < .001, R^2 = .70 (Table 5c). Positive Communality, b = 0.93, SE = 0.10, t = 8.94, p < .001, Agency, b = 0.52, SE = 0.11, t = 4.73, p < .001, and Ideology, b = 0.62, SE = 0.09, t = 6.54, p < .001, effects emerged.

Like Study 1, a Communion × Ideology interaction emerged, b = -0.23, SE = 0.09, t = -2.52, p = .01. Communality effects were stronger for more conservative, b = 1.15, SE = 0.11, t = 10.65, p < .001, than liberal, b = 0.70, SE = 0.16, t = 4.37, p < .001, participants. Unexpectedly, Description did not qualify this interaction, b = -0.07, SE = 0.09, t = 0.74, p = .46.

The Agency × Communion × Ideology interaction emerged, b = -0.21, SE = 0.08, t = 2.61, p = .01 (Figure 4c). When Harris was perceived as less agentic, similar communality effects emerged among more conservative, b = 0.95, SE = 0.14, t = 6.97, p < .001, and liberal, b

= 0.91, *SE* = 0.22, *t* = 4.25, *p* < .001, participants, *b* = 0.04, *SE* = 0.24, *t* = 0.16, *p* = .87. When more agentic, communality effects were stronger for more conservative, *b* = 1.36, *SE* = 0.17, *t* = 8.21, *p* < 0.001, than liberal, *b* = 0.49, *SE* = 0.19, *t* = 2.63, *p* = .01, participants, *b* = 0.86, *SE* = 0.24, *t* = 3.62, *p* = .001.

Voting Likelihood

The model was significant, $F(15, 247) = 44.15$, $p < .001$, $R^2 = .73$ (Table 5d). Positive Communality, *b* = 1.20, *SE* = 0.12, *t* = 9.77, *p* < .001. Agency, *b* = 0.51, *SE* = 0.13, *t* = 3.96, *p* < .001, and Ideology, *b* = 1.15, *SE* = 0.11, *t* = 10.39, *p* < .001, effects emerged.

The expected Communion × Ideology × Description interaction emerged, *b* = -0.26, *SE* = 0.11, *t* = 2.49, *p* = .01 (Figure 3c). With the agentic description, communality effects similarly emerged across more conservative, *b* = 1.25, *SE* = 0.16, *t* = 8.10, *p* < .001, and liberal, *b* = 1.61, *SE* = 0.27, *t* = 6.06, *p* < .001, participants, *b* = -0.36, *SE* = 0.27, *t* = 1.35, *p* = 0.18. With the communal description, communality effects were stronger for more conservative, *b* = 1.30, *SE* = 0.20, *t* = 6.41, *p* < 0.001, than liberal, *b* = 0.62, *SE* = 0.27, *t* = 2.28, *p* = .02, participants, *b* = 0.69, *SE* = 0.33, *t* = 2.11, *p* = 0.04.

The Agency × Communion × Ideology interaction emerged, *b* = -0.45, *SE* = 0.09, *t* = 4.84, *p* < .001 (Figure 4d). When Harris was perceived as less agentic, communality effects were stronger for more liberal, *b* = 1.48, *SE* = 0.25, *t* = 5.82, *p* < .001, than more conservative, *b* = 0.73, *SE* = 0.16, *t* = 4.57, participants, *b* = -0.74, *SE* = 0.28, *t* = -2.62, *p* = .01. When more agentic, communality effects were stronger for more conservative, *b* = 1.82, *SE* = 0.20, *t* = 9.34, *p* < .001, than liberal, *b* = 0.75, *SE* = 0.22, *t* = 3.42, *p* = .001, participants, *b* = 1.07, *SE* = 0.28, *t* = 3.80, *p* = .001.

Table 5: Regression Summaries for Evaluations of Kamala Harris by Description, Gendered Trait Inferences and Political Ideology in Study 2 (N = 263).

	Impression Positivity			**Support**			**Expected Success**			**Voting Likelihood**		
	b (SE)	*T*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*
Communality	1.30 (0.08)	13.42	< .001	1.31 (0.12)	10.53	< .001	0.93 (0.10)	8.94	< .001	1.20 (0.12)	9.77	< .001
Agency	0.22 (.10)	2.14	.03	0.45 (0.13)	3.42	< .001	0.52 (0.11)	4.73	< .001	0.51 (0.13)	3.96	< .001
Ideology	0.65 (0.09)	7.36	< .001	1.07 (0.11)	9.49	< .001	0.62 (0.09)	6.55	< .001	1.15 (0.11)	10.39	< .001
Description	-0.06 (0.08)	0.68	.50	-0.02 (0.11)	0.16	.88	-0.13 (0.09)	1.40	.16	0.002 (0.11)	0.03	.98
Communality x Agency	0.14 (0.08)	1.86	.06	-0.02 (0.10)	.22	.82	-0.003 (0.08)	0.04	.97	0.09 (0.10)	0.95	.35
Communality x Ideology	-0.19 (0.08)	2.29	.02	-0.08 (.11)	0.79	.43	-0.23 (0.09)	2.52	.01	-0.09 (0.11)	0.77	.44
Agency x Ideology	-0.12 (0.10)	1.19	.23	-0.21 (0.13)	1.64	.10	-0.02 (0.11)	0.21	.83	-0.30 (0.13)	2.32	.02
Communality x Description	-0.15 (0.09)	1.51	.13	-0.24 (0.12)	1.89	.06	-0.07 (0.10)	0.66	.51	-0.24 (0.12)	1.94	.05
Agency x Description	0.06 (0.10)	0.63	.53	0.10 (0.13)	0.78	.43	-0.04 (0.11)	0.37	.72	0.09 (0.13)	0.69	.49
Ideology x Description	0.02 (0.09)	0.20	.84	0.06 (0.11)	0.50	.62	0.12 (0.09)	1.24	.21	0.13 (0.11)	1.21	.23

Communality x Agency x Ideology	-0.20 (0.07)	2.71	.007	-0.43 (0.09)	4.48	< .001	-0.21 (0.08)	2.61	0.01	-0.45 (0.09)	4.84	< .001
Communality x Agency x Description	0.09 (0.08)	1.15	.25	0.002 (0.10)	0.03	.98	-0.04 (0.08)	0.47	.64	0.06 (0.10)	0.67	.50
Communality x Ideology x Description	-0.20 (0.08)	2.45	.02	-0.29 (0.11)	2.69	.008	-0.07 (0.09)	0.74	.46	-0.26 (0.11)	2.49	.01
Agency x Ideology x Description	0.11 (0.10)	1.11	.27	0.13 (0.13)	1.01	.31	0.15 (0.11)	1.36	.18	0.10 (0.13)	0.80	.42
Communality x Agency x Ideology x Description	-0.01 (0.07)	0.17	.86	-0.02 (0.09)	0.24	.81	-0.003 (0.08)	0.03	.97	-0.03 (0.09)	0.34	.74

Interpreting Political Ideology Effects

Regression models again showed that a more liberal political ideology more strongly related to favorable evaluations than the other variables (Supplemental Material (A) Table A2).

Discussion

The extent to which Harris was perceived as more agentic than communal was larger when people read a description of Harris conveying agentic relative to communal roles and behaviors. This pattern extends work on trait inferences of female politicians (Bauer & Carpinella, 2018; Conroy et al., 2020; Conroy & Green, 2020; Gervais & Hillard, 2011) by suggesting malleable inferences of female candidates of color based on how they are described. This finding may be important given that many stereotypes of women of color do not include traits characteristic of communality (Rosette et al., 2016).

When the description conveyed more agentic details, positive communality effects on impression positivity, support, and voting likelihood similarly emerged across perceivers. This finding aligns with theoretical (Haines & Stroessner, 2019b) and empirical (Cassidy & Liebenow, 2021; Heilman & Okimoto, 2007; Rosette et al., 2016; Rosette & Tost, 2010b; Schneider et al., 2022) work suggesting that perceived communality is broadly important for agentic women to bypass agentic penalties. By contrast, stronger positive communality effects emerged among more conservative than more liberal perceivers when the description conveyed more communal details. A female candidate's communality may thus offset negative effects of perceiving that candidate as less communal if people share her ideology. Indeed, across these evaluations, more liberal than conservative perceivers evaluated Harris positively to a greater extent when she was perceived as less communal. Providing different information about female

candidates may thus affect the extent that sharing that candidate's ideology allows for more positive evaluations.

The patterns for expected success differed from the other evaluations. Like Study 1, more liberal relative to conservative perceivers expected success to a greater extent when perceiving Harris as less communal. Whereas the other evaluations regarded *perceiver-focused* attitudes and behaviors, expected success regarded how *Harris would act* in the context of a leadership role. Although beyond the scope of the current work, these patterns suggest that ideology and perceived communality may interact to affect evaluations differently based on whether evaluations focus on the perceiver or candidate. A more liberal ideology, for example may offset perceived low communality when people evaluate how women with shared ideology might perform in office because people show favoritism toward ideologically similar others in terms of their job performance (West & Iyengar, 2020). Notably, people may admit that a female candidate is likely to do well without explicitly supporting her (e.g., Cassidy & Krendl, 2019). Because this possibility holds implications for political polling, future work should directly address it.

Perceived agency qualified an interaction between perceived communality and ideology across evaluations. When Harris was perceived as more agentic, positive communality effects were stronger among more conservative than more liberal participants. When perceived as less agentic, similarly positive communality effects on impression positivity and expected success emerged across partisans. In fact, *stronger* positive communality effects on support and voting likelihood emerged among more liberal than more conservative perceivers. Notably, trait inferences across studies were quite positive. Having more positive inferences of Harris on one trait dimension might offset more negative evaluations on the other provided that perceivers

shared an ideology with her and thus were motivated to like her (e.g., Mallinas et al., 2018). Indeed, when people perceived Harris as less agentic, more liberal relative to more conservative perceivers who perceived her as more communal were more likely to vote for her. Likewise, when people perceived Harris as more agentic, more liberal relative to more conservative perceivers who perceived her as less communal more positively evaluated her.

One question regards why this interaction emerged in Study 2, but not Study 1. Given the subtleties of three-way interactions, it could be difficult to detect in the smaller sample used in Study 1. Because subtle effects can yield sizable consequences, future work should probe such relations while powering for a complicated interaction. We also made a methodological choice to have people make trait inferences before the other evaluations to highlight their contributions to those evaluations. Doing so may have allowed subtle effects of the inferences to emerge. That this interaction emerged further supports that political ideology and gendered trait inferences interacted to affect evaluations of Harris. Showing different patterns between and across different descriptions of Harris suggests multiple ways these interactive effects emerged to affect evaluations.

General Discussion

This work identified effects of gendered trait inferences and political ideology on evaluations of Kamala Harris, the first woman of color elected to the executive branch. Harris was perceived as more agentic than communal (Studies 1 and 2) even when a description of Harris conveyed her communality (Study 2). Across studies, perceiving Harris as more communal and having a more liberal political ideology each positively related to evaluations of her. People with a more liberal relative to a more conservative political ideology had weaker

(albeit significant) positive communality effects when evaluating her expected success (Studies 1 and 2) and when her communal roles and behaviors were described (Study 2).

Aspiring female leaders receive backlash (e.g., Rudman & Glick, 1999) for deviating from the prescription that women should be communal (Prentice & Carranza, 2002a), evidenced by, for example, lower political support (Bauer, 2017; Bauer & Carpinella, 2018). However, agentic women who are perceived as communal can receive evaluative advantages (Cassidy & Liebenow, 2021; Rosette & Tost, 2010). Aligning with these findings, people who perceived Harris as more communal had more positive evaluations of her despite her being perceived as more agentic overall. This pattern may reflect that the prioritization of agentic *and* communal behavior prevents negative bias against women (Haines & Stroessner, 2019b). Being perceived as having agentic and communal behaviors and goals may benefit female politicians the most in terms of garnering support. Indeed, violating communal prescriptions rather than more general norm deviations explains backlash against female candidates (Okimoto & Brescoll, 2010).

As a woman of color, Harris is subject to racial scrutiny (Dowe, 2020) and stereotypes (Rosette et al., 2016) that have not applied to past female candidates for United States executive office. Some work suggests that perceived communality may not strongly contribute to evaluations of agentic women who people identify as Black (Livingston et al., 2012) because stereotypes of Black women largely do not involve traits characteristic of communality (Rosette et al., 2016). Yet, stereotypes of Black women also contribute to their worse leadership assessments relative to White women (Motro et al., 2021). In actuality, communality may be key for women of color to overcome penalties afforded by their agency. That communality strongly contributed to evaluations of Harris across studies supports the latter idea. Perceived communality may be even *more* important for women of color (relative to White women)

aspiring to high leadership roles to be positively evaluated by others. Future work should directly address this possibility.

Perceiver Ideology may Qualify Communality Effects on Evaluations

Positive communality effects on evaluations of Harris were weaker for more liberal than more conservative people when a description of her conveyed details about her communality (but not agency). Despite the current polarized political climate (Finkel et al., 2020), this finding suggests that female candidates cannot assume unwavering support from people sharing their ideology. It also suggests a context by which co-partisan support of agentic women may be more likely. When evidence of Harris's communality was presented, liberal relative to conservative perceivers more positively evaluated Harris when their perceptions of Harris were lower in communality. Because people favor candidates sharing their ideologies (Iyengar & Krupenkin, 2018; Mallinas et al., 2018; West & Iyengar, 2020), a shared aspect of identity may allow people to favorably evaluate agentic female candidates provided they have some evidence of communality.

Although we interpreted interactive effects of gendered trait inferences and political ideology on evaluations of Harris from an identity-based perspective, this interpretation does not rule out effects of sexism on evaluations of agentic women aspiring to leadership roles. Indeed, work on Hillary Clinton's presidential campaign has indicated that sexism was related to her lack of support (Cassidy & Krendl, 2019; Ratliff et al., 2019; Rothwell et al., 2019). Although Clinton and Harris both had historic nominations, they differed subtly in how they impacted the status quo of men in power. Clinton's nomination shattered a glass ceiling for women because she was positioned to be the President of the United States. Harris's election certainly shattered a glass ceiling because she became the first woman elected to the executive branch. However, endorsing

Harris maintained the male-dominated presidency. One avenue for future research may consider how factors contributing to evaluations of a woman poised to *be* President relative to *aid* a President may be different. For example, prejudice might more strongly affect evaluations of candidates of color when a party ticket cannot help maintain a status quo of a White man in power. Indeed, racial prejudice negatively related to support for Barack Obama in 2008 (e.g., Payne et al., 2010; Piston, 2010) and 2012 (e.g., Knuckey & Kim, 2015).

Future Directions

Few women attain the success that Harris has had during her political career. It will thus be important that future work examine whether the displayed patterns generalize to other women. Examining this generalizability is important because Democrat and Republican female candidates may present themselves differently to their constituents. Hillary Clinton and Sarah Palin, for example, were differentially perceived in their agency and communality in 2008, and whereas only perceived communality positively related to a likelihood of voting for Clinton, both perceived agency and communality did for Palin (Gervais & Hillard, 2011). Since these patterns were for idiosyncratic candidates, it is unclear whether differences emerged because of candidate qualities or because of broader partisan differences. Future work may systematically vary the presentation of hypothetical conservative and liberal female candidates.

Such work is also important to conduct given that people associate Democrats more with femininity and Republicans more with masculinity (Winter, 2010). Because Harris is a Democrat, communal traits associated with femininity may be especially important for her to be perceived as being aligned with her party. This finding does not suggest, however, that Republican women will not face any penalties for their agency. Indeed, more women run for office and win as Democrats than Republicans (Winter, 2010). One possibility is that female

Republican candidates are more likely penalized given the positive relation between conservatism and traditional gender role beliefs. Thus, it will be important to determine not just *if* the current findings seem to generalize, but *why* they seem to do so to capture the likely considerable nuances contributing to evaluations of female political candidates.

It would also be useful for work to consider how evaluations of female candidates change over time. The current data were collected when support for Harris may have been most polarized among partisans (i.e., in the weeks preceding the 2020 election). Perceptions of female politicians, however, change over time (e.g., Cassidy & Krendl, 2019). Thus, future work may examine the described relations at different time points to reveal whether they are stable or whether they change in response to different events (e.g., party losses in midterm elections). Such work may broaden our understanding how people evaluate agentic women in ecologically valid contexts.

Practical Implications

Despite more woman being interested in political office than ever before (Bonneau & Kanthak, 2020), women remain underrepresented in the United States government. The current work stresses a consideration of gendered trait inferences when developing women's campaign strategies and suggests that this consideration may be especially relevant for female candidates of color. Highlighting multiple prioritized aspects of their lives may be key for women to continue shattering glass ceilings in government representation. Women may thus choose to be strategic in how they present themselves to the electorate in different contexts. Some people may also assume that women do not suffer disproportionately based on their gender in terms of garnering public support. The current work argues against this potential assumption by providing consistent evidence that the gendered ways in which people perceive an actual female candidate

affects their evaluations of her, sometimes despite a shared ideology. More awareness of the influence of gendered trait inferences on evaluations of female candidates may provide a timely reminder to the electorate that we do not live in a post-gender or post-racial society.

Conclusion

The findings of our study indicate that perceived communality is key for agentic women to garner support (e.g., Haines & Stroessner, 2019). We found consistently strong positive communality effects on evaluations of Kamala Harris and showed that political ideology tempered the strength of this relation in some contexts. By identifying when and how gendered trait inferences affected the support of an actual female candidate, we can more fully understand contributing factors to women's underrepresentation in politics.

.

CHAPTER III: TRAIT INFERENCES FROM THE "BIG TWO" PRODUCE GENDERED EXPECTATIONS OF FACIAL FEATURES

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.

Introduction

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 et al., 2020) have focused on observations that such stereotypes are tied to traditional beliefs about social roles (Diekman & 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 et al., 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 et al., 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.

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 et al., 2015) and behave (Hehman et al., 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 et al., 2021). Although people largely agree on face impressions, such impressions do not strongly relate to actual behaviors (e.g., Rule et al., 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.

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 (Diekman & 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 hirable 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.

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 et al., 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 et al., 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 femininized 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 masculinize faces receive fewer votes (Hehman et al., 2014a). 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 femininized image representations of language arts than physics teachers' faces (Degner et al., 2019), which is consistent with people's perceptions of STEM fields as non-communal (Brown et al., 2015). Moreover, some work (e.g., Degner et al., 2019; Imhoff et al., 2013) has assessed how people mentally represent the "big two" in faces (e.g.,

Oliveira et al., 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.

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 et al., 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; L. A. Rudman et al., 2012) and through tendencies by men to score higher than

women on measures linked to traditional social role endorsement (Eagly & Karau, 2002; Fiske et al., 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.

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 valanced 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.

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.

Method

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.

Stimuli

To measure expectations of sexually dimorphic facial features, we created stimuli following procedures from past work (Cassidy & Krendl, 2018; Cole et al., 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 et al., 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 Material (B) 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 (B) for more information on viewing conditions.

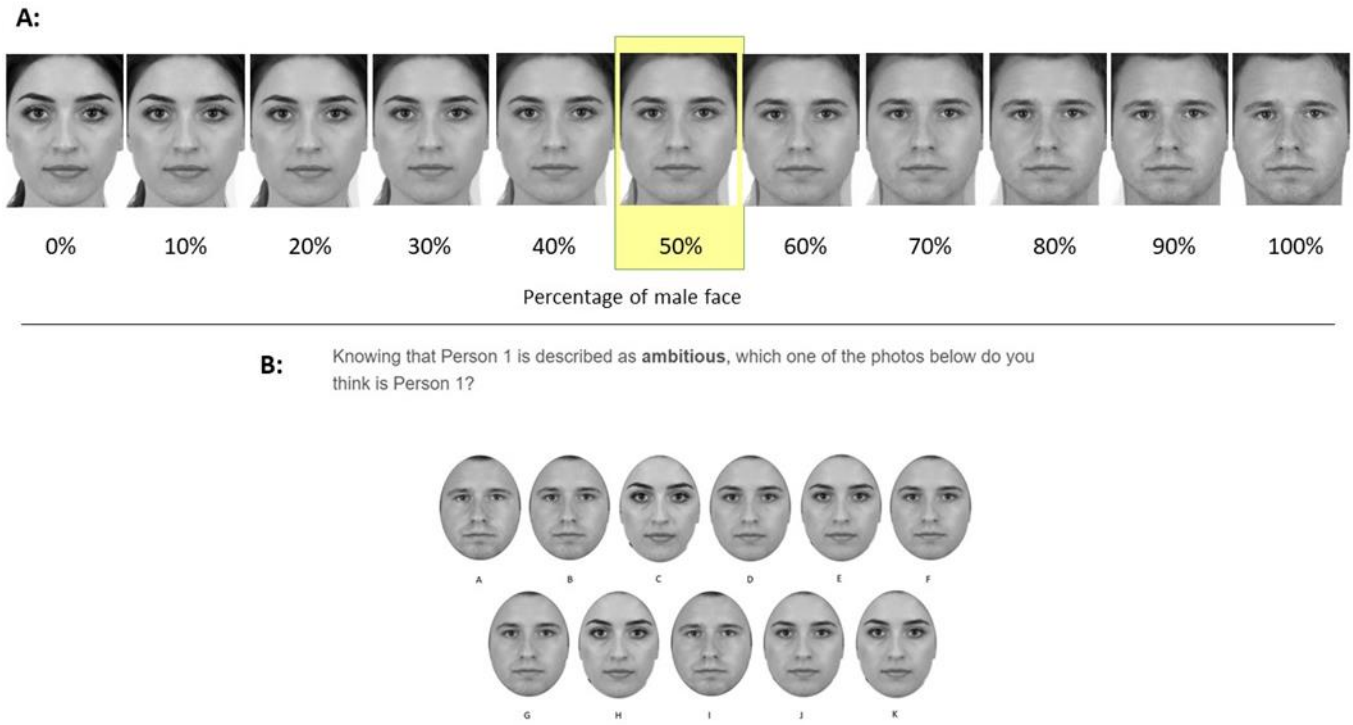
Procedure

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 et al., 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

(Figure 5b). Participants selected which face best represented the target six times within each block.

Figure 5: Female-male Face Morph Process



Note. 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).

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.

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.

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 $\alpha = .96$) were averaged to create composite political ideology scores ($M = 5.83$, $SD = 2.37$).

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 strongly*) to 5 (*agree strongly*). Higher scores indicate more sexism. Items measuring hostile (Cronbach's $\alpha = .93$; $M = 1.46$, $SD = 1.14$) and benevolent (Cronbach's $\alpha = .92$; $M = 1.90$, $SD = 1.21$) sexism were averaged to create hostile and benevolent sexism scores.

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 (Cronbach's $\alpha = .90$; $M = 29.07$, $SD = 20.46$).

People lastly provided demographic information.

Results

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 6a for descriptive and inferential statistics.

Table 6: 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	.003	0.40 [0.13, 0.66]
Benevolent sexism	2.12 (1.16)	1.63 (1.21)	3.07	.002	0.41 [0.15, 0.68]
Traditional gender role beliefs	32.26 (19.35)	25.15 (21.19)	2.63	.009	0.35 [0.09, 0.62]
Political ideology	5.72 (2.16)	5.97 (2.41)	0.79	.42	0.11 [-0.16, 0.37]
Age	36.99 (10.77)	39.16 (11.62)	1.45	.15	0.19 [-0.07, 0.46]
Years of Education	15.52 (2.47)	15.13 (2.15)	1.24	.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	< .001	0.47 [0.20, 0.75]
Benevolent sexism	2.07 (0.98)	1.62 (1.18)	3.03	.003	0.42 [0.15, 0.70]
Traditional gender role beliefs	31.88 (16.91)	22.79 (18.98)	3.66	< .001	0.51 [0.23, 0.79]
Political ideology	4.61 (1.77)	4.72 (1.86)	-0.45	.66	0.06 [-0.21, 0.34]
Age	37.84 (10.93)	45.67 (13.24)	-4.70	< .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	.10	0.23 [-0.06, 0.66]
Benevolent sexism	1.86 (1.11)	1.70 (1.14)	0.99	.32	0.14 [-0.16, 0.47]
Traditional gender role beliefs	28.03 (19.60)	25.39 (19.12)	0.95	.34	0.14 [-2.83, 8.10]
Political ideology	4.73 (1.64)	4.64 (1.82)	0.39	.69	0.06 [-0.39, 0.58]
Age	39.09 (11.57)	41.21 (12.56)	-1.23	.22	0.18 [-5.50, 1.27]
Years of Education	15.23 (2.16)	15.33 (1.96)	-0.34	.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	<.001	0.49 [0.26, 0.90]
Benevolent sexism	1.97 (1.12)	1.83 (1.06)	1.70	.09	0.23 [0.04, 0.56]

Traditional gender role beliefs	30.87 (19.14)	25.73 (18.98)	1.97	.05	0.27 [-0.01, 10.28]
Political ideology	4.74 (1.59)	4.87 (1.73)	-0.57	.57	0.08 [-0.58, 0.32]
Age	39.45 (11.72)	41.23 (11.89)	-1.10	.27	0.15 [-4.96, 1.41]
Years of Education	15.59 (2.21)	15.43 (2.51)	0.52	.60	0.07 [-0.47, 0.80]
Social dominance orientation	2.35 (1.44)	1.94 (1.30)	2.15	.03	0.30 [0.03, 0.78]
Identity threat	1.89 (1.14)	2.95 (1.40)	-6.08	<.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	<.001	0.51 [0.27, 0.88]
Benevolent sexism	1.97 (1.12)	1.83 (1.06)	0.94	.35	0.13 [-0.15, 0.43]
Traditional gender role beliefs	28.15 (20.06)	26.31 (18.35)	0.70	.49	0.10 [-3.36, 7.03]
Political ideology	4.63 (1.84)	4.59 (1.73)	0.16	.87	0.02 [-0.44, 0.52]
Age	36.48 (9.86)	40.50 (11.46)	-2.70	.01	0.37 [-6.95, 1.09]
Years of Education	15.18 (2.43)	15.36 (2.52)	-0.53	.60	0.07 [-0.86, 0.49]
Social dominance orientation	2.28 (1.41)	2.08 (1.24)	1.11	.27	0.15 [-0.16, 0.56]
Identity threat	1.84 (1.09)	3.06 (1.37)	-7.05	<.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	.90	0.02 [-0.42, 0.48]
Benevolent sexism	2.26 (1.04)	2.56 (1.01)	-1.52	.13	0.29 [-0.69, 0.09]
Traditional gender role beliefs	39.24 (20.39)	39.53 (20.39)	-0.07	.94	0.01 [-7.96, 7.38]
Political ideology	4.69 (1.51)	4.63 (1.79)	0.21	.83	0.04 [-0.54, 0.67]
Age	39.41 (11.13)	38.25 (10.26)	0.57	.57	0.11 [-2.90, 5.23]
Years of Education	15.01 (2.27)	15.70 (2.36)	-1.59	.12	0.30 [-1.56, 0.17]
Social dominance orientation	2.72 (1.49)	2.80 (1.34)	-0.29	.77	0.06 [-0.62, 0.46]

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].

The big two elicited expectations of gendered facial features.

Analytic plan

Across experiments, we fitted mixed-effects models using *lme4* (Bates et al., 2015) in R, estimating confidence intervals using the confint function. P-values were calculated using *lmerTest* (Kuznetsova et al., 2017). Estimated marginal means were obtained using *emmeans* (Lenth, 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.

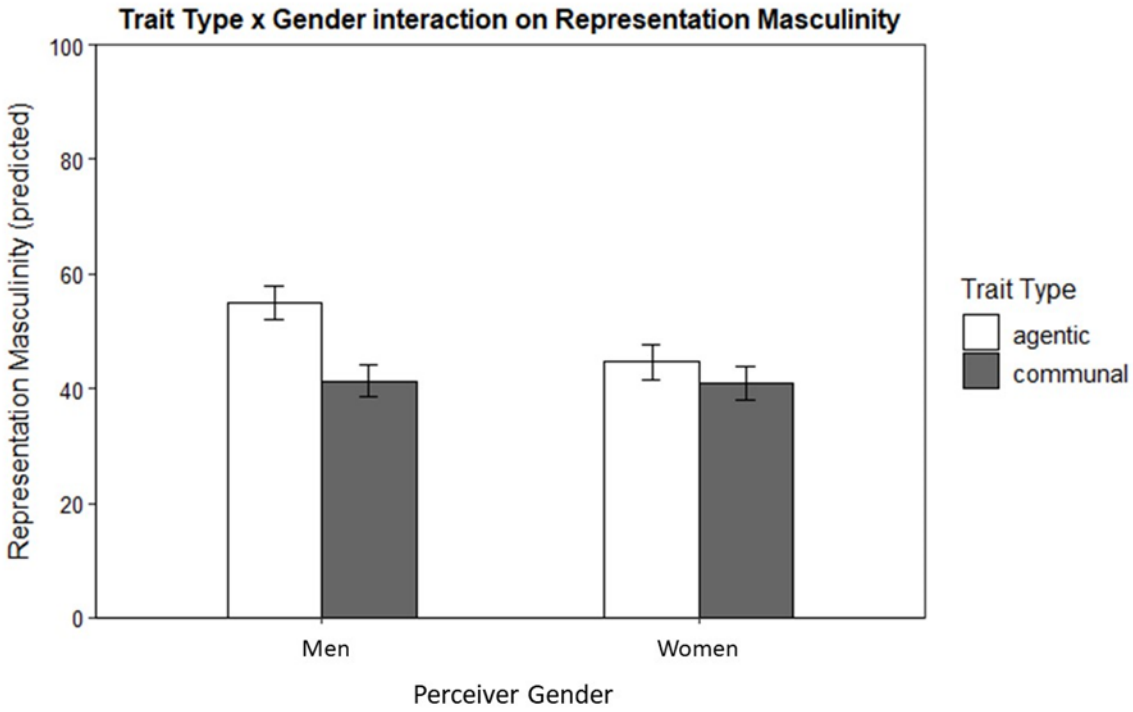
Expectations of gendered facial features

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] (Figure 6). 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, 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 = .80 and alpha = .05. The effect size from the sensitivity analysis is the standardized regression coefficient.

Figure 6: Trait Type by Perceiver Gender Interaction in Experiment 1



Note. 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.

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.

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 femininized 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 et al., 2010; Stanaland et al., 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).

Method

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.

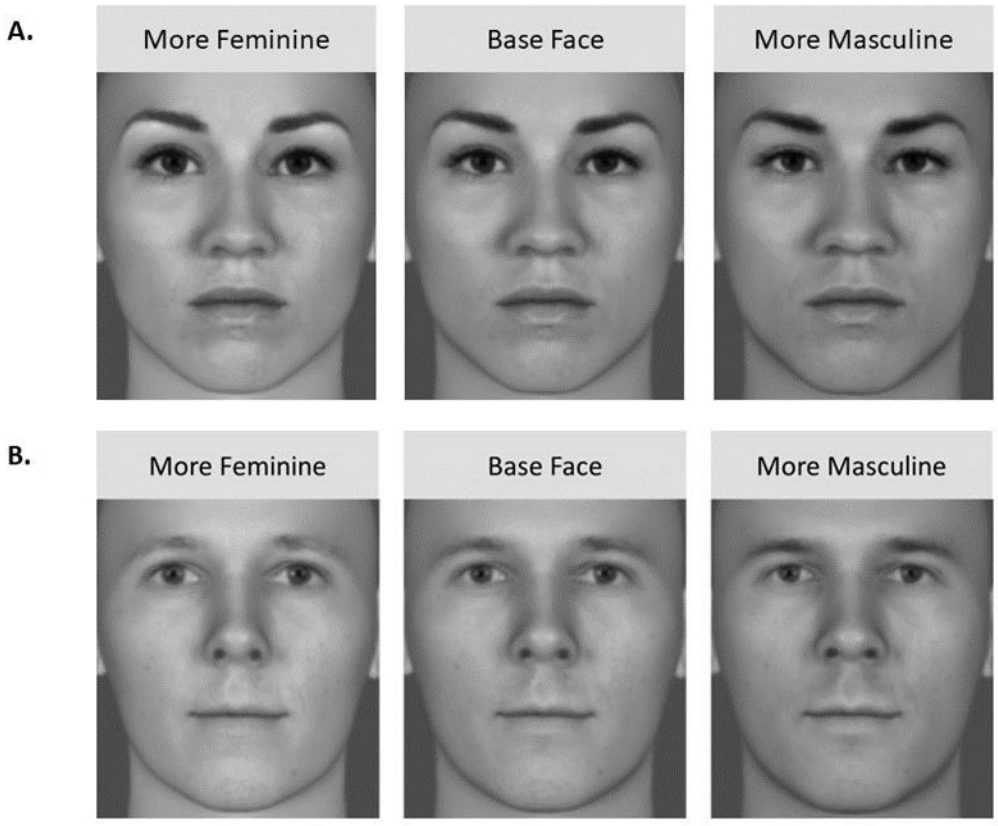
Stimuli

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 (B) 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 (Figure 7). 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, 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 (B)). 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 (B)). 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.

Figure 7: Example Face Images



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

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 et al., 2007; Mitteroecker et al., 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 et al., 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.

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.

Participant Characterization

Participants completed the measures from Experiment 1 in a random order. Responses to political ideology items (Cronbach's α = .96) were averaged to create a composite score (*M* = 4.65, *SD* = 1.81). Items measuring hostile (Cronbach's α = .94; *M* = 1.68, *SD* = 1.29) and benevolent (Cronbach's α = .88; *M* = 1.89, *SD* = 1.08) sexism were averaged to make composite scores. SRQ responses (Cronbach's α = .88) were averaged to make a social role belief score (*M* = 34.25, *SD* = 22.47).

Results

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 6b for descriptive and inferential statistics.

The big two elicited expectations of gendered facial features.

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 7a 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 8a). 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.

Table 7: 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	< .001	-0.25 (0.09)	-2.82	0.0005	-0.82 (0.06)	-12.82	< .001	-0.68 (0.06)	-10.78	< .001
Face Gender	-0.01 (0.04)	-0.16	.87	-0.13 (0.07)	-1.96	0.05	0.02 (0.04)	0.49	.62	-0.005 (0.05)	-0.10	.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	.34	0.07 (0.06)	1.08	.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	.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	.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	.31	-0.03 (0.03)	-1.18	.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	.88	-0.001 (0.02)	-0.06	.95

Table 8: 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 × 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]

Expectations of facial masculinity

Like Experiment 1, the Trait Type effect showed that people selected the masculnized 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 masculnized 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 = .80 and alpha = .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 masculnized female (Estimate = 0.49, 95% CI [0.46, 0.52]) than male (Estimate = 0.53, 95% CI [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 ($\text{Estimate}_{\text{female face}}$ = 0.51, 95% CI [0.47, 0.55]; $\text{Estimate}_{\text{male face}}$ = 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 = 0.01. By contrast, men and women were similarly endorsed masculinized female faces across traits, OR = 0.94, SE = 0.079, z = -0.771, p = .44.

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, 2010b); 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.

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 escribed 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.

Method

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.

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.

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 $\alpha = 0.89$; $M = 26.88$, $SD = 19.39$).

Results

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 6c for descriptive and inferential statistics.

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, $b = 0.17$, $SE = 0.07$, $z = 2.25$, $p = 0.02$, 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, $b = 0.01$, $SE = 0.03$, $z = 0.39$, $p = 0.70$. A sensitivity analysis for this interactive effect indicated a minimum detectable effect of log odds = 0.19 with power = .82 and alpha = .05.

There was also a Face Gender effect, b = -0.13, SE = 0.07, z = -1.96, p = 0.05, 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 7b for other coefficient information and Table 8b for random effects information.

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 valanced 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, 2021a). 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.

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 et al., 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 et al., 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.

Method

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.

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.

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 α = 0.95; M = 4.80, SD = 1.65). Items measuring hostile (Cronbach's α = 0.94; M = 1.49, SD = 1.21) and benevolent (Cronbach's α = 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 α = 0.89; M = 28.45, SD = 19.19).

Social dominance orientation. People completed an eight-item social dominance orientation questionnaire (Ho et al., 2015) to measure their preference for inequality amongst 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 α = 0.92; M = 2.15, SD = 1.39).

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).

Results

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 8a for descriptive and inferential statistics.

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] (Figure 8b). 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 = .82 and alpha = .05. See Table 7c for other coefficient information and Table 8c for random effects information.

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 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 et al., 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.

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.

Method

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.

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.

Participant Characterization

Responses to political ideology items were averaged to create a composite score (Cronbach's α = 0.97; *M* = 4.61, *SD* = 1.78). Items measuring hostile (Cronbach's α = 0.93; *M* = 1.44, *SD* = 1.16) and benevolent (Cronbach's α = 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 α = 0.88); *M* = 27.13, *SD* = 19.11). Social dominance orientation items were averaged to make a composite score (Cronbach's α = 0.92; *M* = 2.17, *SD* = 1.32). Participants indicated gender identity threat (*M* = 2.52, *SD* = 1.39).

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) × 2 (Behavior: analytical, interpersonal) repeated-measures ANOVA. Qualifying a main effect of Trait, $F(1, 212) = 82.78$, $p < .001$, $\eta_p^2 = .28$, was an expected Trait × Behavior interaction, $F(1, 212) = 295.53$, $p < .001$, $\eta_p^2 = .58$. 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, $F(1, 212) = 0.14$, $p = .71$, $\eta_p^2 < .01$.

Results

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 8b for descriptive and inferential statistics.

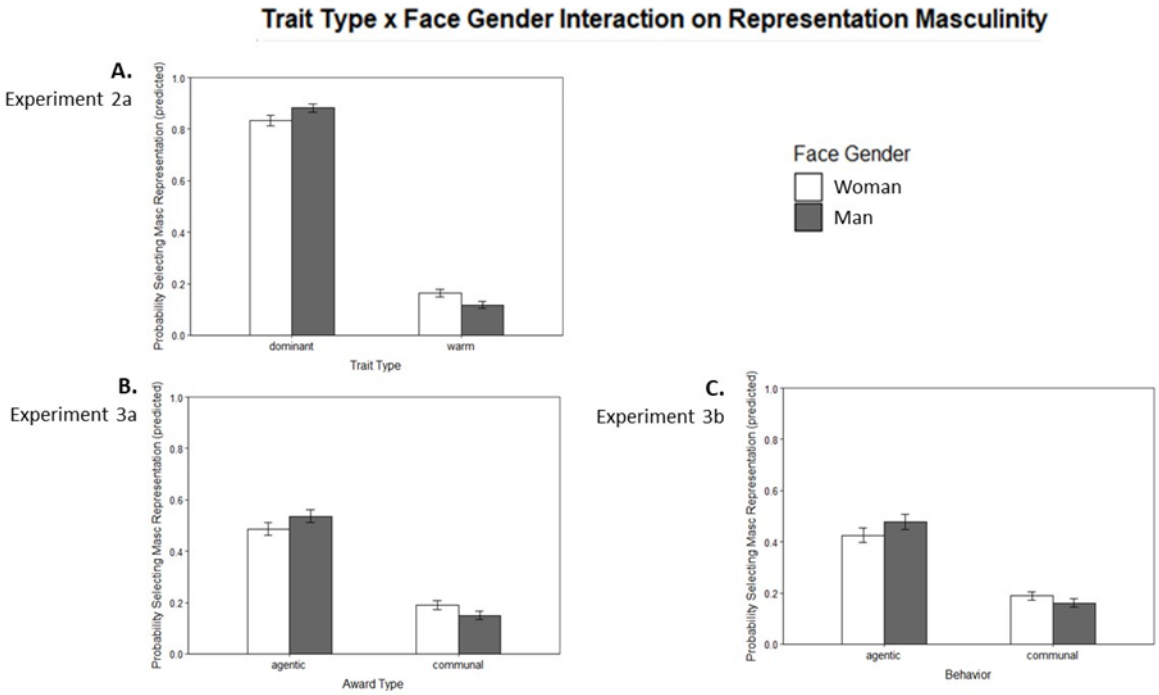
The big two elicited expectations of gendered facial features.

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 < 0.01, 95% CI [0.04, 0.16] (Figure 8c). 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 = .81 and alpha = .05.

Unique to Study 3b was a Trait Type by Perceiver Gender interaction, b = -0.17, SE = 0.06, z = -2.83, p < 0.01, 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 masculnized 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 7d for other coefficient information and Table 8d for random effects information.

Figure 8: Trait Type by Face Gender Interaction (Experiments 2a, 3a, and 3b)



Note. 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.

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 et al., 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 et al., 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.

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 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.

Method

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.

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 = 0.62.

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 et al., 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.

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 α = 0.93; M = 4.67, SD = 1.61). Items measuring hostile (Cronbach's α = 0.92; M =

1.87, *SD* = 1.18) and benevolent (Cronbach's α = 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 α = 0.82); *M* = 39.35, *SD* = 20.30). Social dominance orientation items were averaged to make a composite score (Cronbach's α = 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 Figure 9 for examples). Whereas some work (van Rijsbergen et al., 2014) applies statistical analysis to, for example, the pixel intensities of CIs, assess featural differences, much social perception work (Blais et al., 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 CIs. We used the latter approach here.

Figure 9: Agentic and Communal CIs

Note. Examples of agentic (A & B) and communal (C & D) CIs from Experiment 4.

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 CIs from Phase 1 in a random order and were unaware of what prompt yielded the CIs. 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 CIs reduces the possibility of Type I error inflation in reverse correlation findings (Cone et al., 2021).

Results

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 6 for descriptive and inferential statistics.

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 CIs (119 mental representations; $\alpha = .97$). We also

separately validated high interrater reliability across the agentic (60 mental representations; α = .81) and communal (59 mental representations; α = .88) CIs. 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 (B) 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.

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 et al., 2022; Gingras et al., 2023;

Imhoff et al., 2013; Petsko et al., 2020) and that is detailed on primers on using reverse correlation to examine social perception (Brinkman et al., 2017; Degner et al., 2019; Dotsch et al., 2013; Dotsch & Todorov, 2012; Hess et al., 2023; Imhoff et al., 2013; Oliveira et al., 2019; Oliver et al., 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 et al., 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.

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 et al., 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 important 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 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 et al., 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 et al., 2008, 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 et al., 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 et al., 2012a). 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.

.

CHAPTER IV: MASCULINIZED FACIAL FEATURES NEGATIVELY AFFECT EVALUATIONS OF WOMEN STUDENT LEADERS

Abstract

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

Introduction

Women are underrepresented in leadership in workplace domains ranging from STEM (Kahn & Ginther, 2017) to politics (Baskaran & Hessami, 2018) to academia (Llorens et al., 2021). This underrepresentation undermines women's sense of belonging (Baskaran & Hessami, 2018; Broockman, 2014; Creamer, 2012; Piatek-Jimenez et al., 2018), perpetuating gender disparity in positions traditionally held by men (Gadassi & Gati, 2009; Glick et al., 1995). Much

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

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

& Karau, 2002) asserts women displaying agentic behaviors characteristic of leadership are often perceived as deviating from communal expectations, which elicits negative outcomes for them.

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

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

Stronger impressions of agentic traits from facial features relate to evaluated leadership suitability (Little, 2014; Re, DeBruine, Jones, & Perrett, 2013; Spisak et al., 2012). Inferences of

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

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

A direct link showing sexually dimorphic facial features to pose a barrier to women aspiring to leadership roles, however, has yet to be established. For example, much work examining outcomes for aspiring women leaders varying in sexually dimorphic facial features has not compared these women against aspiring men leaders (e.g., Cassidy & Liebenow, 2021). Thus, we cannot discern whether these features are a barrier unique to women rather than a more general one. Other work has shown that people more negatively evaluate women's (versus men's) faces when they appear gender atypical versus typical (Sutherland et al., 2015). However, this work has done so without determining whether these evaluations extend to differential outcomes for men and women. Without these outcomes, we cannot discern whether patterns from backlash research (Moss-Racusin, 2014; Rudman & Glick, 1999; Rudman & Phelan, 2008) are reflected in evaluations of aspiring leaders varying in sexually dimorphic facial features.

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

Indirectly supporting this premise, recent work showed that STEM professors with more trustworthy (versus dominant) faces are perceived as valuing communal goals and as affording

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

Second, we focus on targets *before* they obtain leadership positions versus targets already enjoying them. Although affordances from STEM professor appearance may signal a lab's communal or agentic culture (Joshi et al., 2023), these affordances are made of targets who have already achieved a leadership role. Focusing on targets before leadership is attained is important to test whether sexually dimorphic features may function as a nonverbal *barrier* to leadership rather than a set of features that elicit inferences from established leaders. Notably, feminine appearance can (inaccurately) signal that women (but not men) are not well-suited for STEM faculty positions (Banchefsky et al., 2016). Such work suggests that sexually dimorphic facial

features have the potential to invalidate women's relative to men's leadership aspirations to a greater extent.

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

Experiment 1

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

more negative outcomes based on sexually dimorphic facial features only for women aspiring to leadership.

Method

Participants

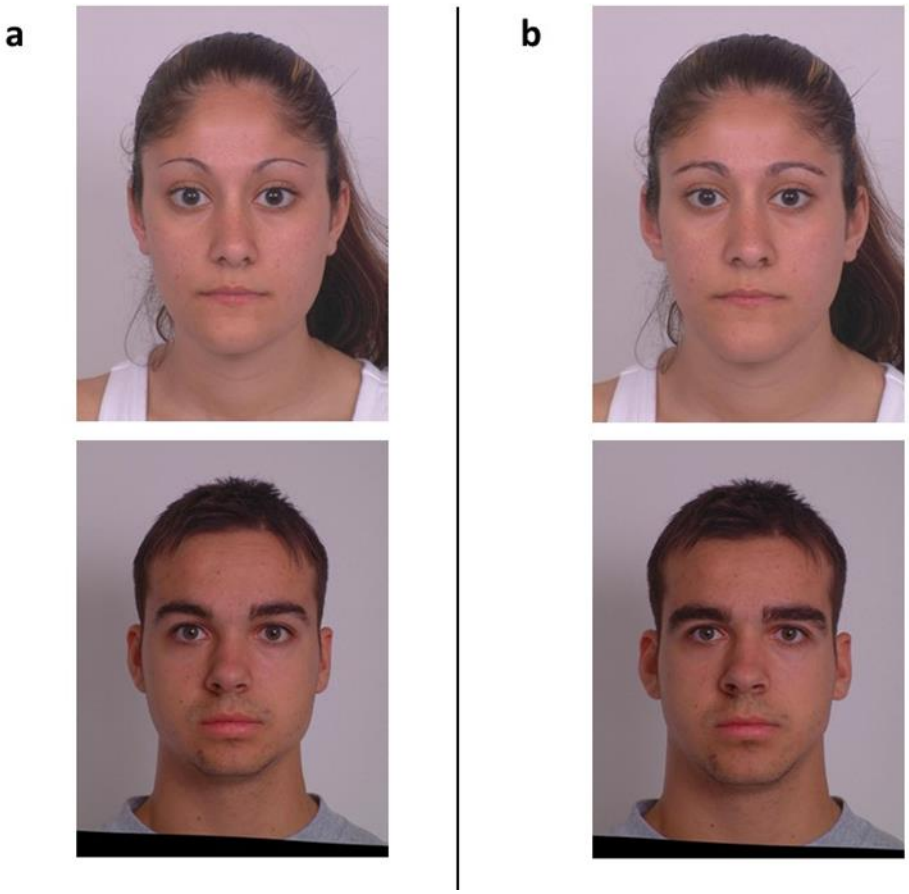
Based on recent work (e.g., Wen et al., 2020), we recruited 200 MTurk participants. Twenty-seven failed at least one attention check (described below), yielding 173 analyzed participants (M_{age} = 39.61 years, SD = 11.88; $M_{years\ of\ education}$ = 15.33, SD = 2.34; 69 identifying as female). Of these participants, 135 identified as White, 17 as Black, 14 as Asian, four as multi-racial, two as American Indian/Alaska Native, and one as Native Hawaiian/other Pacific Islander. Of these participants, 157 identified as non-Hispanic. All experiments were approved by the UNC Greensboro IRB. All participants provided informed consent. To benefit future work, we provide sensitivity analyses for interactive effects between sexually dimorphic facial features and target gender on outcomes across experiments (see below). Data and analytic code are available on the Open Science Framework (https://osf.io/ygwj5/?view_only=2ad8762941c84303a2a5a969b5629aef).

Stimuli

Twenty identities each of neutrally expressive White men and White women used in work examining contributions of sexually dimorphic facial features to social outcomes (e.g., Goh et al., 2021) were selected. Each face identity had a masculinized and feminized version (Figure 10; for details on face manipulation, see DeBruine & Jones, 2017) that have been used in related work (e.g., DeBruine & Jones, 2017; Little et al., 2011; Shiramizu et al., 2024). Moreover, using the same identities across task versions manipulates sexually dimorphic features while holding all other face aspects constant. Masculinized and feminized faces broadly differed in similar

ways. Masculinized faces had thicker, lower eyebrows and wider, square jaws. Feminized faces had thinner, higher eyebrows and smaller, round jaws. Relevant here, the faces were all emerging adults and thus appropriate for evaluation as potential college-aged student leaders.

Figure 10: Example Feminized and Masculinized Face Identities



Note. Example feminized (a) and masculinized (b) face identities.

Task

Participants evaluated candidates, saw candidate pictures, and provided impressions. The description read, "The Student Policies Manager is a prestigious position for a student leader who displays strong competency in their work, independence, and motivation to lead their college community. Responsibilities for this position are to advise the board of trustees on

student and university issues, lead student council in weekly meetings, and be a competent facilitator of student and faculty queries. To succeed in this role, the candidate should be highly competent, independent, intelligent, and determined. Past leadership experience is desirable. Past students in this position have gone onto work for prestigious companies and even been elected for local, city, and state legislative seats." A one-tailed t-test against the scale (1 [*very feminine*] to 7 [*very masculine*]) midpoint showed that 40 participants who did not complete the experiments evaluated the description as gender neutral (M = 3.90, SD = 0.22), $t(38) = -0.51$, $p = .61$, $d = 0.082$.

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

In the main task, participants saw pictures of and evaluated 20 White men and 20 White women in randomly presented trials. Within each gender, ten faces each were masculinized or feminized. Each identity was shown once and was displayed through all evaluations. Two task versions counterbalanced target facial features within-participants. In each self-paced trial, participants rated a candidate on **warmth, competence,** and **dominance** using scales ranging from 1 (*not at all*) to 7 (*extremely*). These traits are commonly used in work assessing agency (competence and dominance) and communion (warmth; (Abele et al., 2016; Fiske et al., 2007; Wen et al., 2020). Critically, they have been extensively linked to gender stereotypes (Fiske et al., 2002c; Imhoff et al., 2013) and used in related work on impressions from sexually dimorphic

facial features (Walker & Wänke, 2017a). The General Discussion includes a discussion of potential limitations of using these traits.

On the next screen, participants evaluated the candidate ("How likely are you to endorse this candidate for the role of Student Policies Manager?"; "How likely is this candidate to do well as Student Policies Manager?"; "How likely are other people to like reporting to this candidate as the Student Policies Manager?") using scales ranging from 1 (*not at all likely*) to 7 (*extremely likely*).

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

Characterization Measures

We characterized participants on measures collected in a random order (see Table 9 for descriptive statistics and intercorrelations). Participants indicated political ideology over four items (overall, economic issues, social issues, and foreign policy issues) on a scale ranging from 1 (*extremely conservative*) to 9 (*extremely liberal*). Responses (Cronbach's $\alpha = 0.95$) were averaged. Participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). The ASI measures hostile (e.g., "Women are too easily offended") and benevolent (e.g., "Women should be cherished and protected by men") sexism on a scale ranging from 0 (*disagree strongly*) to 5 (*agree strongly*). Higher scores indicate more sexism. Items measuring hostile (Cronbach's $\alpha = .91$) and benevolent (Cronbach's $\alpha = .89$) sexism were averaged. Participants completed the Social Roles Questionnaire (SRQ; Baber & Tucker, 2006) to measure traditional social role beliefs. Participants indicated how much they agreed with each statement (e.g., "For many important jobs, it is better to choose men instead of women") from 0% to 100% in 10% increments. Higher scores reflect more traditional social role beliefs. Items (Cronbach's $\alpha = .90$)

were averaged. Participants completed a social dominance orientation questionnaire (Ho et al., 2015) using a scale ranging from 1 (*strongly oppose*) to 7 (*strongly favor*). We averaged responses (e.g., "An ideal society requires some groups to be on top and others to be on the bottom."; Cronbach's $\alpha = 0.91$).

Table 9: Intercorrelations for participant characterization measures

(Means, standard deviations, and correlations with confidence intervals) for Experiments 1 (below diagonal) and 2 (above diagonal).

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

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

Data Analyses

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

Results

Gender Differences on Characterization Measures

Men and women participants had similar hostile sexism, benevolent sexism, traditional social role beliefs, social dominance orientation, political ideology, age, and years of education (Table 10a).

Effects of Gender and Sexually Dimorphic Facial Features on Trait Evaluations

We regressed evaluations on Trait (effects coded with competence as the reference level), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions. We included Perceiver Gender on an exploratory basis. See Table 11 for all coefficient information for the fixed effects. The random effects structure allowed intercepts to vary by participant and face identity and facial characteristics effects to vary by participant and face identity.

Table 10: Descriptive (*M* [*SD*]) and inferential statistics for participant characterization measures in Experiments 1 (a) and 2 (b)

a. Experiment 1	**Men (*N* = 104)**	**Women (*N* = 69)**	***t***	***p***	***d* [95% CI]**
Hostile sexism	1.87 (1.17)	1.63 (1.22)	1.33	.18	0.21 [-0.12, 0.61]
Benevolent sexism	2.05 (1.10)	2.09 (1.17)	-0.28	.78	0.04 [-0.39, 0.30]

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

Table 11: Effects of Face Gender, Facial Characteristics, and Participant Gender on Evaluated Warmth (a), Competence (b), and Dominance (c) in Experiment 1.

	a. Warmth			b. Competence			c. Dominance		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*p*
(Intercept)	4.21	4.03 – 4.39	**<0.001**	4.53	4.36 – 4.69	**<0.001**	4.39	4.23 – 4.56	**<0.001**
Face Gender	0.26	0.13 – 0.38	**<0.001**	0.23	0.12 – 0.34	**<0.001**	-0.04	-0.14 – 0.05	0.381
Facial Characteristics	0.12	0.08 – 0.17	**<0.001**	0.07	0.03 – 0.10	**<0.001**	-0.11	-0.15 – -0.07	**<0.001**
Participant Gender	0.10	-0.03 – 0.23	0.127	0.09	-0.03 – 0.21	0.156	0.16	0.03 – 0.29	**0.017**
Face Gender * Facial Characteristics	0.03	-0.01 – 0.07	0.147	0.04	0.00 – 0.07	**0.025**	0.11	0.08 – 0.14	**<0.001**
Face Gender * Participant Gender	-0.07	-0.09 – -0.04	**<0.001**	-0.04	-0.07 – -0.02	**0.002**	0.02	-0.01 – 0.05	0.143
Facial Characteristics * Participant Gender	-0.02	-0.05 – 0.01	0.178	-0.01	-0.04 – 0.02	0.588	0.02	-0.02 – 0.05	0.386

Note. Participant Gender: (-1 = Men participants, 1 = Women participants), Face Gender: (-1 = Men's faces, 1 = Women's faces), Facial Characteristics: (-1 = Masculine, 1 = Feminine).

Main Effects

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

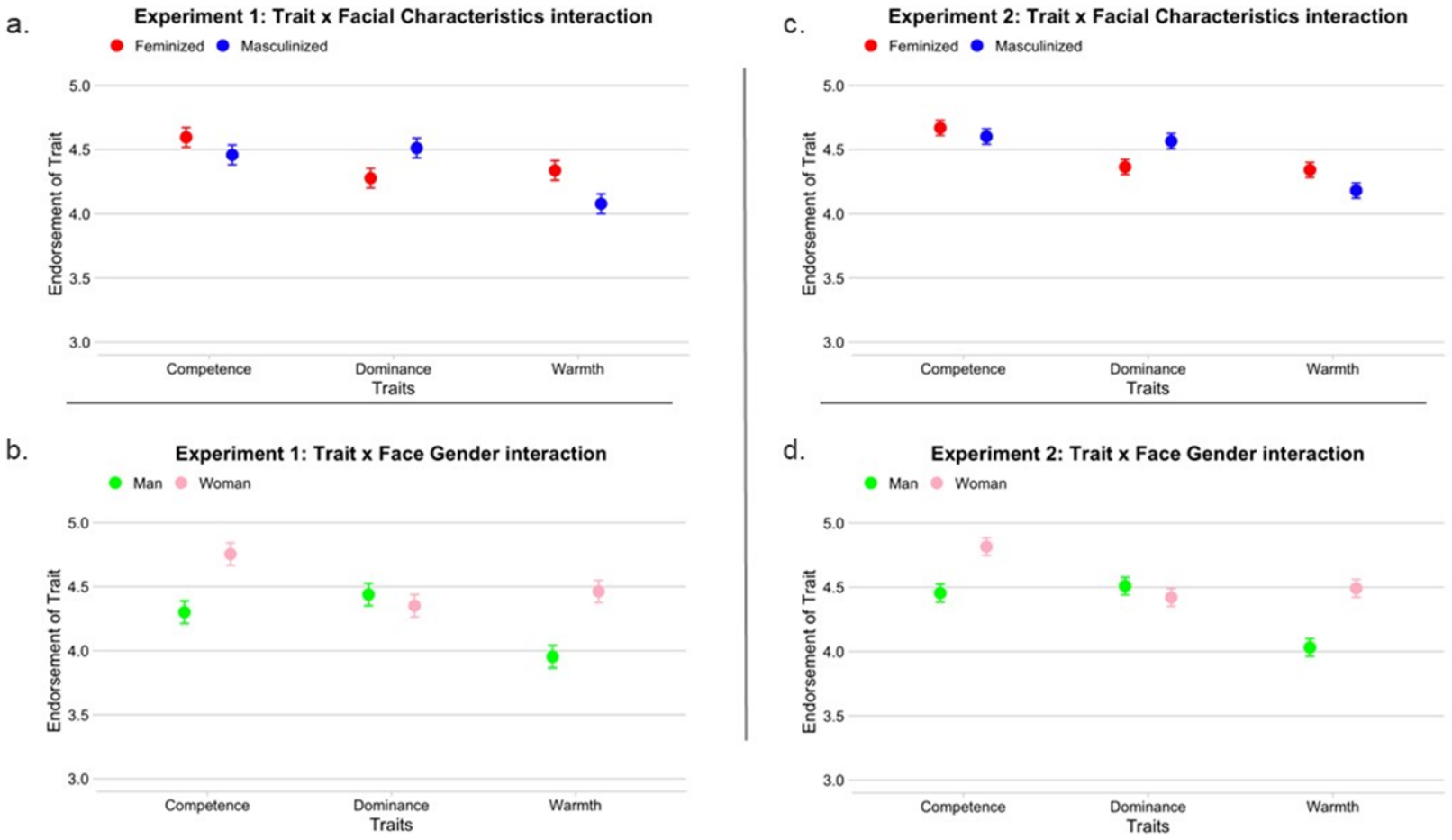
Two-Way Interactions

Trait and Face Gender interacted (Figure 11a). Participants evaluated women's (Estimate = 4.46, 95% CI [4.29, 4.63]) relative to men's (Estimate = 3.95, 95% CI [3.78, 4.12), faces as more warm, $b = 0.51$, SE = 0.09, $z = 5.62$, $p < 0.01$. Participants evaluated women's (Estimate = 4.75, 95% CI [4.58, 4.93]) relative to men's (Estimate = 4.30, 95% CI [4.13, 4.47]), faces as more competent, $b = 0.45$, SE = 0.09, $z = 4.99$, $p < .001$. In contrast, participants evaluated women's (Estimate = 4.35, 95% CI [4.18, 4.52]) and men's (Estimate = 4.44, 95% CI [4.27, 4.61]) faces similarly on dominance, $b = -0.09$, SE = 0.09, $z = -0.96$, $p = 0.93$.

Trait and Facial Characteristics interacted (Figure 11b). Participants evaluated feminized (Estimate = 4.34, 95% CI [4.19, 4.49]) relative to masculinized (Estimate = 4.08, 95% CI [3.93, 4.23), faces as more warm, $b = 0.26$, SE = 0.04, $z = 7.35$, $p < .001$. Participants evaluated feminized (Estimate = 4.60, 95% CI [4.45, 4.75]) relative to masculinized (Estimate = 4.46, 95%

CI [4.31, 4.61]), faces as more competent, b = 0.14, SE = 0.04, z = 3.89, p = .001. In contrast, participants evaluated femininized (Estimate = 4.28, 95% CI [4.13, 4.43]) relative to masculinized (Estimate = 4.51, 95% CI [4.36, 4.66]), faces as less dominant, b = -0.23, SE = 0.04, z = -6.65, p < .001.

Figure 11: Trait by Facial Characteristics and Trait by Face Gender Interactions Across Experiments 1 and 2

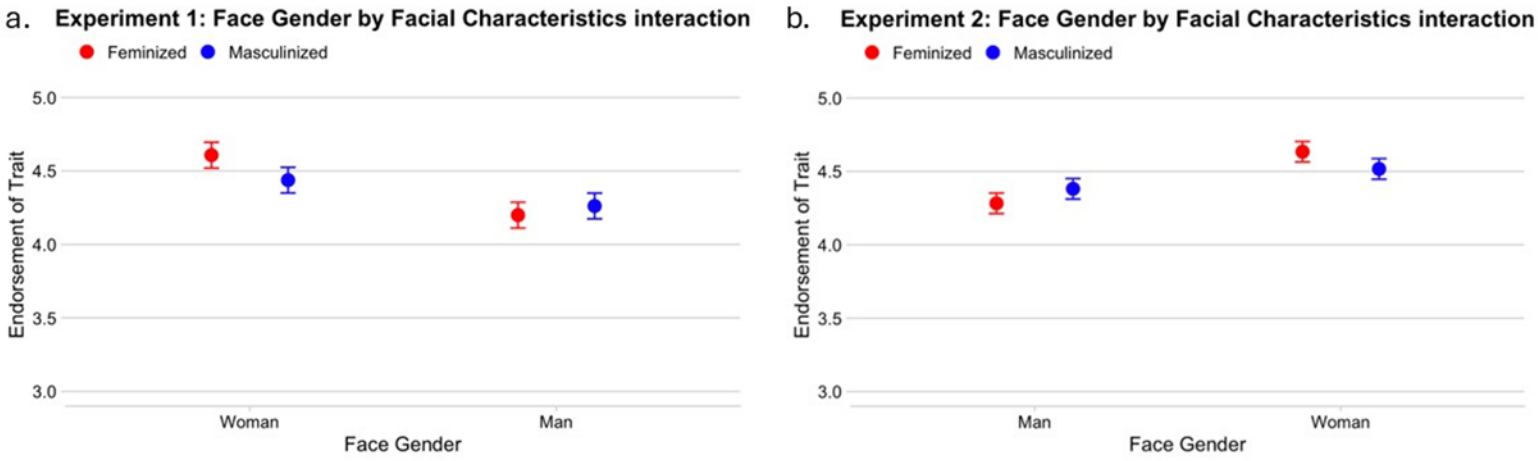


Note. Across Experiments 1 (a) and 2 (c), interactions between trait and facial characteristics emerged on trait endorsements. Participants had stronger trait endorsements of competence and warmth for feminized versus masculinized faces, while stronger trait endorsements of dominance for masculinized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

Facial Characteristics and Face Gender interacted (Figure 12a). Participants evaluated feminized versus (Estimate = 4.61, 95% CI [4.44, 4.78]) masculinized (Estimate = 4.44, 95% CI [4.27, 4.61]) women's faces higher in traits overall, $b = 0.17$, SE = 0.03, $z = 4.99$, $p < .001$. In contrast, participants similarly evaluated traits in feminized (Estimate = 4.20, 95% CI [4.03, 4.37) versus masculinized (Estimate = 4.26, 95% CI [4.09, 4.43]) men's faces, $b = -0.06$, SE = 0.03, $z = -1.82$, $p = 0.26$.

Face Gender and Perceiver Gender interacted. Men evaluated women's (Estimate = 4.43, 95% CI [4.24, 4.62]) versus men's (Estimate = 4.08, 95% CI [3.89, 4.28]), faces as higher in traits, $b = 0.34$, SE = 0.09, $z = 3.93$, $p < .001$. To a lesser extent, women evaluated women's (Estimate = 4.61, 95% CI [4.39, 4.83]) versus men's (Estimate = 4.38, 95% CI [4.16, 4.60), faces as higher in Traits, $b = 0.24$, SE = 0.09, $z = 2.62$, $p = .04$.

Figure 12: Face Gender by Facial Characteristics Interactions on Trait Endorsement Across Experiments 1 and 2



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

Three-Way Interactions

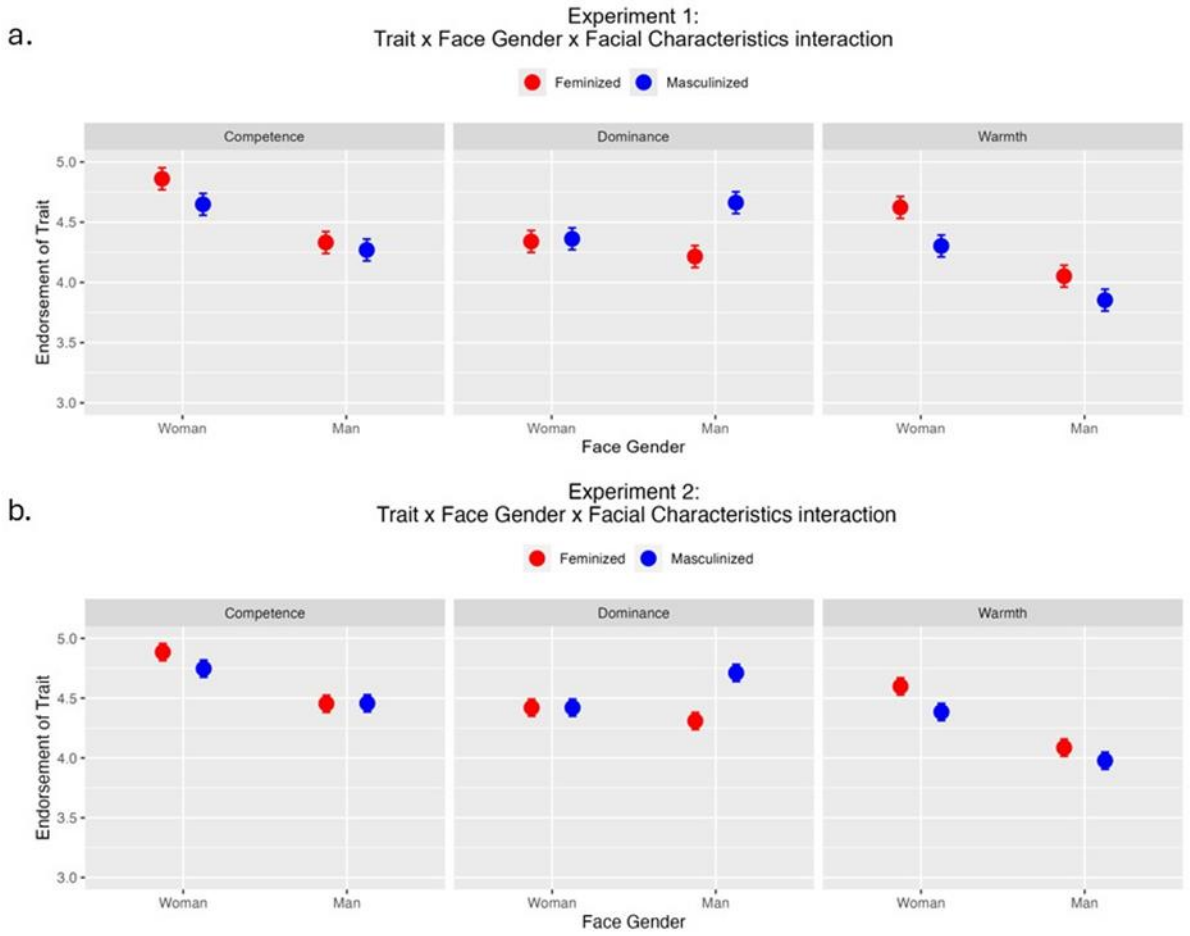
Trait, Facial Characteristics, and Face Gender. Trait qualified the interaction between Facial Characteristics and Face Gender (Figure 13a). We characterized this interaction by examining the Facial Characteristics effect for each Face Gender at each Trait.

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

For women candidates, participants evaluated feminized (Estimate = 4.86, 95% CI [4.68, 5.04]) versus masculinized (Estimate = 4.65, 95% CI [4.47, 4.83]), faces as more competent, b = 0.21, SE = 0.05, z = 4.42, p < .001. For men candidates, in contrast, participants evaluated feminized (Estimate = 4.33, 95% CI [4.15, 4.51]) versus masculinized (Estimate = 4.27, 95% CI [4.09, 4.45]) faces as similarly competent, b = 0.06, SE = 0.05, z = 4.14, p = .98.

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

Figure 13: Trait by Face Gender by Facial Characteristics Interactions Across Experiments 1 and 2



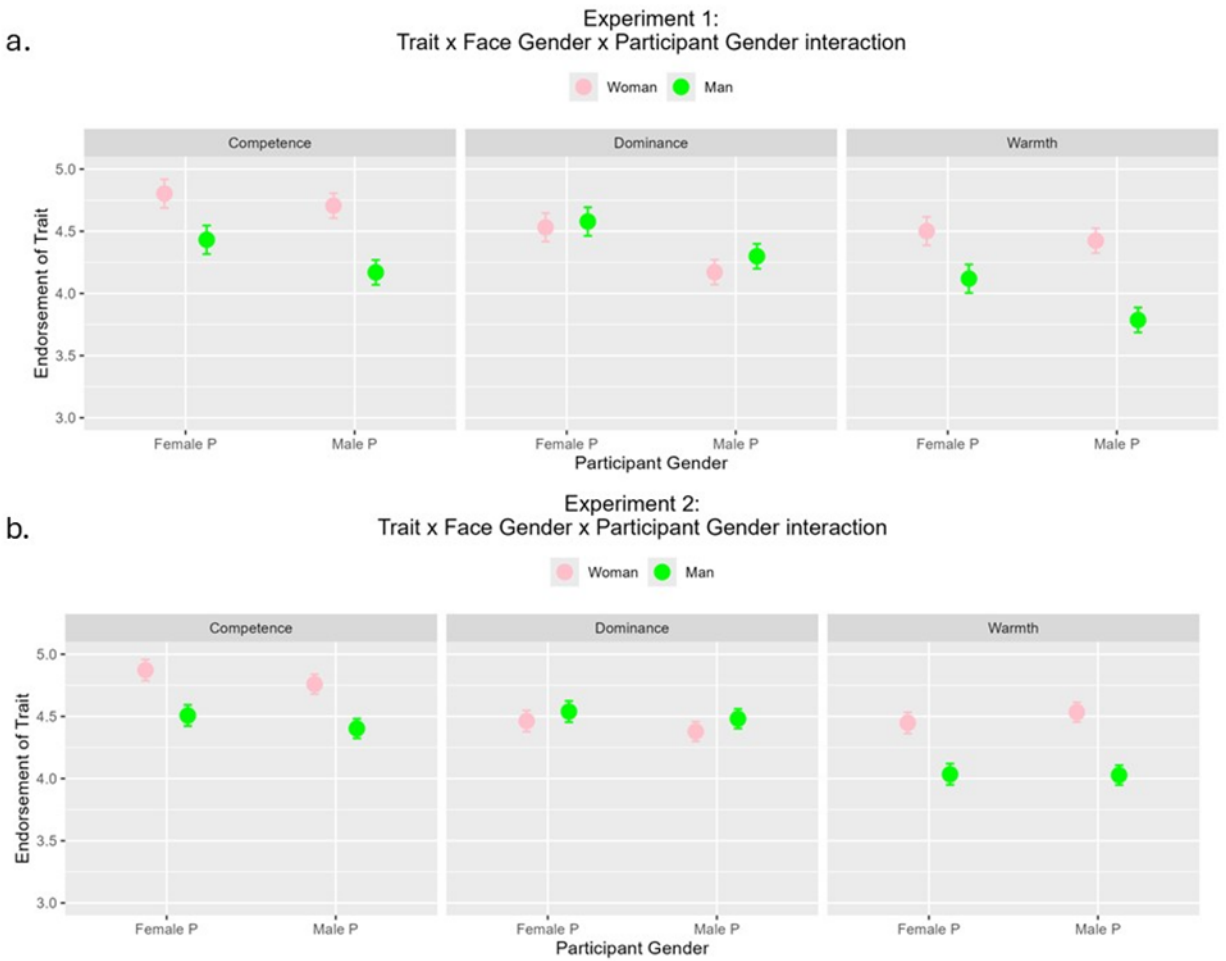
Trait, Face Gender, and Perceiver Gender. Trait qualified the interaction between Face Gender and Perceiver Gender (Figure 14a). We characterized this interaction by examining how evaluations of Traits differed for each Face and Perceiver Gender.

Men evaluated men's faces as more dominant (Estimate = 4.30, 95% CI [4.10, 4.49]) than competent (Estimate = 4.17, 95% CI [3.97, 4.37]), *b* = -0.13, *SE* = 0.04, *z* = -3.515, *p* = 0.02. In contrast, they evaluated women's faces as more competent (Estimate = 4.71, 95% CI [4.51, 4.90]) than dominant (Estimate = 4.17, 95% CI [3.97, 4.37]), *b* = 0.54, *SE* = 0.04, *z* = 14.49, *p* < .001. Men evaluated men's faces as more competent (Estimate = 4.17, 95% CI [3.97, 4.37]) than arm (Estimate = 3.79, 95% CI [3.59, 3.98]), *b* = 0.38, *SE* = 0.04, *z* = 10.36, *p* <.001. They also evaluated women's faces as more competent (Estimate = 4.71, 95% CI [4.51, 4.90]) than warm (Estimate = 4.42, 95% CI [4.23, 4.62]), *b* = 0.28, *SE* = 0.04, *z* = 7.63, *p* <.001. Men evaluated men's faces as more dominant (Estimate = 4.30, 95% CI [4.10, 4.49]) than warm (Estimate = 3.79, 95% CI [3.59, 3.98]), *b* = 0.51, SE = 0.04, *z* = 13.88, *p* <.001. In contrast, they rated women's faces as more warm (Estimate = 4.42, 95% CI [4.23, 4.62]) than dominant (Estimate = 4.17, 95% CI [3.97, 4.37]), b = -0.25, SE = 0.04, z = -6.86, p <.001.

Unlike men, women similarly evaluated men's faces in dominance (Estimate = 4.58, 95% CI [4.35, 4.80]) and competence (Estimate = 4.43, 95% CI [4.21, 4.66]), *b* = -0.15, *SE* = 0.05, *z* = -3.23, *p* = 0.06. Like men, women evaluated women's faces as more competent (Estimate = 4.80, 95% CI [3.97, 4.37]) than dominant (Estimate = 4.53, 95% CI [4.31, 4.76]), *b* = 0.27, *SE* = 0.05, *z* = 5.99, *p* <.001. Like men, women evaluated men's faces as more competent (Estimate = 4.43, 95% CI [4.21, 4.66]) than warm (Estimate = 4.12, 95% CI [3.89, 4.34]), *b* = 0.31, *SE* = 0.05, *z* = 6.91, *p* <.001. Also, like men, women evaluated women's faces as more competent (Estimate = 4.80, 95% CI [3.97, 4.37]) than warm (Estimate = 4.50, 95% CI [4.28, 4.73]), *b* = 0.30, *SE* =

0.05, $z = 6.65$, $p < .001$. Like men, women evaluated men's faces as more dominant (Estimate = 4.58, 95% CI [4.35, 4.80]) than warm (Estimate = 4.12, 95% CI [3.89, 4.34]), $b = 0.46$, $SE = 0.05$, $z = 10.13$, $p < .001$. Unlike men, women rated women's faces as similarly warm (Estimate = 4.50, 95% CI [4.28, 4.73]) and dominant (Estimate = 4.53, 95% CI [4.31, 4.76]), $b = 0.03$, $SE = 0.05$, $z = 0.66$, $p > .99$.

Figure 14: Trait by Face Gender by Participant Gender Interactions Across Experiments 1 and 2



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

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

Because responses to the three outcome evaluations were reliable (Cronbach's $\alpha = 0.93$), we created a composite outcome rating by averaging across outcomes. We regressed these composite ratings on Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions. Perceiver Gender was included on an exploratory basis. The model included a random effects structure that allowed intercepts to vary by participant and face identity and that allowed Facial Characteristics effects to vary by participant and face identity. See Table 12 for all model coefficient information. See Supplemental Material (C) for breakdown of each individual outcome variable.

Table 12: Effects of Face Gender, Facial Characteristics, and Participant Gender on (a) Candidate Endorsement, (b) Likelihood of Doing Well in Position, and (c) People Liking Candidate as a Leader in Experiment 1.

	a. Outcome 1			**b. Outcome 2**			**c. Outcome 3**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*p*
(Intercept)	4.11	3.92 – 4.29	**<0.001**	4.26	4.09 – 4.43	**<0.001**	4.22	4.04 – 4.39	**<0.001**
Face Gender	0.26	0.14 – 0.38	**<0.001**	0.25	0.13 – 0.36	**<0.001**	0.20	0.09 – 0.31	**<0.001**
Facial Characteristics	0.07	0.03 – 0.11	**<0.001**	0.07	0.03 – 0.10	**<0.001**	0.07	0.03 – 0.11	**0.001**
Participant Gender	0.12	-0.03 – 0.26	0.108	0.12	-0.00 – 0.25	0.054	0.14	0.00 – 0.27	**0.045**
Face Gender * Facial Characteristics	0.05	0.01 – 0.08	**0.016**	0.03	0.00 – 0.07	**0.045**	0.07	0.03 – 0.11	**<0.001**
Face Gender * Participant Gender	-0.04	-0.07 – -0.01	**0.005**	-0.04	-0.07 – -0.01	**0.006**	-0.06	-0.09 – -0.03	**<0.001**
Facial Characteristics * Participant Gender	-0.01	-0.04 – 0.01	0.329	-0.00	-0.03 – 0.03	0.995	-0.02	-0.05 – 0.01	0.318

Note. Participant Gender: (-1 = Men participants, 1 = Women participants), Face Gender: (-1 = Men's faces, 1 = Women's faces), Facial Characteristics: (-1 = Masculine, 1 = Feminine).

Main Effects

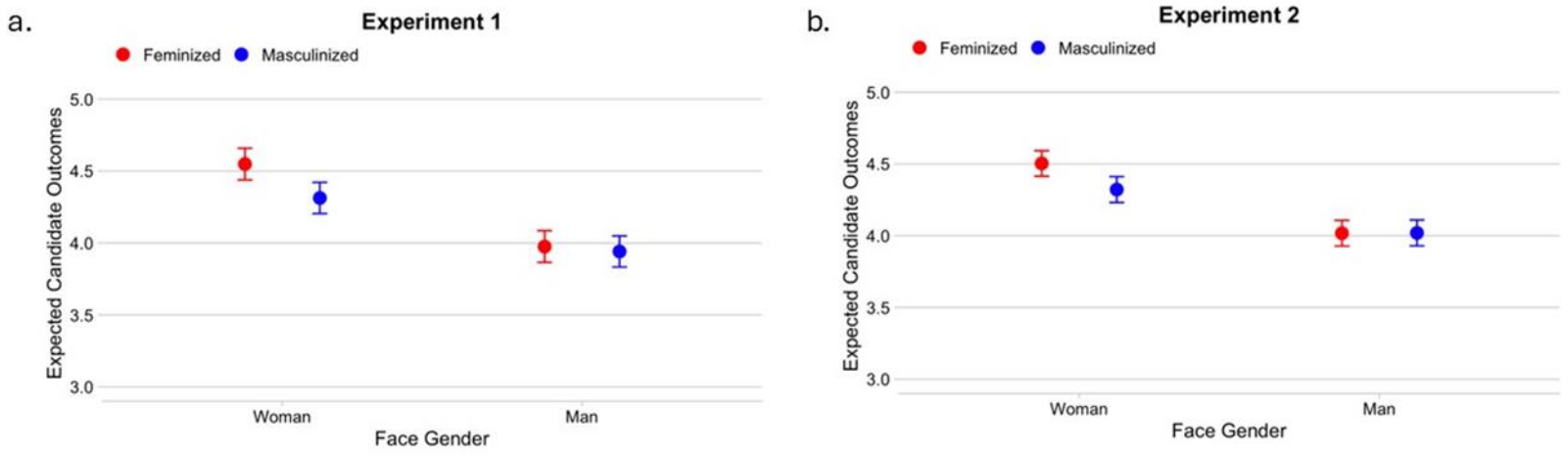
A Face Gender effect showed that women's (Estimate = 4.43, 95% CI [4.22, 4.64]) versus men's (Estimate = 3.96, 95% CI [3.75, 4.17]) faces were endorsed higher in outcomes. A Facial Characteristics effect showed that feminized (Estimate = 4.26, 95% CI [4.09, 4.44]) versus masculinized (Estimate = 4.13,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

Two-Way Interactions

As hypothesized, Face Gender and Facial Characteristics interacted (Figure 15a). Participants endorsed feminized (Estimate = 4.55, 95% CI [4.33, 4.76]) more than masculinized (Estimate = 4.31, 95% CI [4.10, 4.52]) women's faces, $b = 0.24$, SE = 0.05, $z = 4.74$, $p < .001$. No difference between feminized (Estimate = 3.98, 95% CI [3.76, 4.19]) versus masculinized (Estimate = 3.94, 95% CI [3.73, 4.15) men's faces emerged, $b = 0.03$, SE = 0.05, $z = 0.68$, $p = 0.90$. Using our specified model and sample size, a sensitivity analyses revealed a minimum detectable interaction effect size of 0.0605 with 80% power and alpha = .05.

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

Figure 15: Face Gender by Facial Characteristics Interactions on Evaluative Outcomes Across Experiments 1 and 2



Note. Across Experiments 1 (a) and 2 (b), interactions between face gender and facial characteristics emerged on evaluative outcomes. Participants had more positive evaluative outcomes for women candidates when they had feminized versus masculinized facial characteristics.

Discussion

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

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

Evaluations both complemented and contrasted backlash research. In complement (e.g., Rudman, 1998), only women with masculinized relative to feminized faces were less favorably

evaluated. This pattern supports sexually dimorphic facial features as a route to backlash against women. It also suggests an effect emergent in settings involving college-aged students, an understudied population in backlash research.

Women (versus men) candidates and feminized (versus masculinized) faces each received more favorable outcomes. In fact, feminized women's faces received the most favorable outcomes when compared to all other gender and facial characteristics combinations (see Supplemental Material (C)). These patterns may seem surprising given work showing that women versus men are less favorably evaluated when aspiring to leadership (Eagly & Karau, 2002a). Leadership is traditionally aligned with traits better reflecting masculinity (e.g., power) than femininity (e.g., warmth; (Koenig et al., 2011; Rosette & Tost, 2010). Indeed, people are quicker to associate agentic behaviors with leadership when the targets in question are men versus women (Scott & Brown, 2006). Here, women were evaluated as agentic via their higher competence, but were not penalized overall relative to men.

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

leadership positions where a leader is perceived to have less versus more autonomy in decision-making (Smith et al., 2007).

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

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

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

Experiment 2

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

Method

Participants

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

Task

Experiment 2 replicated Experiment 1 with the following changes. Participants were randomly assigned to read an analytical (*n* = 199) or an interpersonal (*n*= 174) description. Men and women participants were evenly distributed across versions, $\chi 2\ (1) < .001$, $p = 0.99$.

The analytical description read, "The Student Policies Manager is a prestigious position for a student leader who displays a strong work ethic, determination, and persistence to lead their college community. Responsibilities for this position are to manage the board of trustees on student and university issues, oversee the student council in weekly meetings, and be highly competent when facilitating queries from students and faculty. To succeed in this role, the candidate should be decisive, self-sufficient, outspoken, and authoritative. Past leadership experience is desirable. Past students in this position have gone onto work for prestigious companies and even been elected for local, city, and state legislative seats."

The interpersonal description read, "The Student Policies Manager is a prestigious position for a student leader who displays a strong community-oriented approach to their work, inclusivity, and cooperation to lead their college community. Responsibilities for this position

are to support the board of trustees on student and university issues, guide the student council in weekly meetings, and create a communal environment to facilitate student and faculty queries. To succeed in this role, the candidate should be nurturing, empathetic, responsive, and trustworthy. Past leadership experience is desirable. Past students in this position have gone onto work for prestigious companies and even been elected for local, city, and state legislative seats."

Forty participants who did not complete the experiments rated descriptions on their interpersonal and analytical nature, valence, and gender stereotypicality using scales from 1 (*not at all*) to 7 (*very*). The analytical job was more than moderately analytical (M = 5.67, SD = 1.52), $t(38) = 7.98$, $p < .001$, $d = 1.28$. The interpersonal job was more than moderately interpersonal (M = 5.92, SD = 1.44), $t(38) = 8.34$, $p < .001$, $d = 1.34$. The interpersonal versus analytical job was more interpersonal, $t(38) = 4.62$, $p < .001$, $d = 1.07$. The analytical versus interpersonal job was more analytical, $t(38) = -4.32$, $p < .001$, $d = 0.91$, The interpersonal (M = 5.31, SD = 1.42) versus analytical (M = 3.46, SD = 1.74) job was more stereotypically feminine, $t(38) = 4.35$, $p < .001$, Cohen's $d = 1.16$. The interpersonal (M = 5.54, SD = 1.27) and analytical (M = 5.13, SD = 1.54) jobs had similar valance, $t(38) = 1.65$, $p = 0.11$, $d = 0.29$.

Four versions counterbalanced job description (analytical/interpersonal) between-participants and facial features (masculinized/feminized) within-participants. Participants indicated following directions (M = 6.82, SD = 0.47).

Beliefs about College Students

After the main task, participants indicated beliefs about men and women college students' warmth (M_{women} = 5.14, SD_{women} = 1.13; M_{men} = 4.37, SD_{men} = 1.43), competence (M_{women} = 5.45, SD_{women} = 1.24; M_{men} = 5.04, SD_{men} = 1.24), dominance (M_{women} = 4.36, SD_{women} = 1.32; M_{men} = 5.28, SD_{men} = 1.20), and responsibility (M_{women} = 5.31, SD_{women} = 1.35; M_{men} = 4.34,

SD_{men} = 1.62) on scales from 1 (*not at all [trait]*) to 7 (*completely [trait]*). Traits were randomly presented.

Participant Characterization

We averaged responses to the political ideology items (Cronbach's $\alpha = 0.95$), the hostile sexism items (Cronbach's $\alpha = .92$), the benevolent sexism items (Cronbach's $\alpha = .88$), the social role belief items (Cronbach's $\alpha = .90$), and the social dominance orientation items. (Cronbach's $\alpha = 0.91$). See Table 9 for intercorrelations and descriptive statistics.

Results

Gender Differences on the Characterization Measures

Men had more hostile sexism, benevolent sexism, and traditional social role beliefs than women. Men and women were similar in social dominance orientation, political ideology, age, and years of education. See Table 10b.

Effects of Gender and Sexually Dimorphic Facial Characteristics on Trait Evaluations

We regressed evaluations on Job Description (-1 = analytical, 1 = interpersonal), Trait (effects coded with competence as the reference level), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions. Perceiver Gender was included on an exploratory basis. See Table 13 for all coefficient information for the fixed effects. The random effects structure allowed intercepts to vary by participant and face identity and facial characteristics effects to vary by participant and face identity.

Main Effects

Main effects of Trait and Face Gender replicated Experiment 1. The Trait effect showed faces rated higher in competence (Estimate = 4.64, 95% CI [4.52, 4.75]) versus dominance

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

Two-Way Interactions

The interaction between Trait and Face Gender replicated Experiment 1 (Figure 11c). Participants evaluated women's (Estimate = 4.49, 95% CI [4.36, 4.63]) versus men's (Estimate = 4.03, 95% CI [3.90, 4.17), faces as more warm, b = 0.46, SE = 0.08, z = 6.02, p < 0.001. Participants evaluated women's (Estimate = 4.82, 95% CI [4.68, 4.95]) versus men's (Estimate = 4.46, 95% CI [4.32, 4.59]), faces as more competent, b = 0.36, SE = 0.08, z = 4.73, p < .001. They similarly evaluated women's (Estimate = 4.42, 95% CI [4.29, 4.56]) and men's (Estimate = 4.51, 95% CI [4.37, 4.65]) dominance, b = -0.09, SE = 0.08, z = -1.17, p = 0.09.

The interaction between Trait and Facial Characteristics partially replicated Experiment 1 (Figure 11d). Like Experiment 1, participants evaluated feminized (Estimate = 4.34, 95% CI [4.23, 4.46]) versus masculinized (Estimate = 4.18, 95% CI [4.06, 4.30), faces as more warm, b = 0.16, SE = 0.03, z = 6.02, p < 0.001. Also like Experiment 1, they evaluated feminized (Estimate = 4.36, 95% CI [4.25, 4.48]) versus masculinized (Estimate = 4.57, 95% CI [4.45, 4.68]), faces as less dominant, b = -0.20, SE = 0.03, z = -7.52, p < .001. Contrasting Experiment 1, they evaluated feminized (Estimate = 4.67, 95% CI [4.55, 4.78]) and masculinized (Estimate = 4.60, 95% CI [4.49, 4.72]) faces as similarly competent, b = 0.07, SE = 0.03, z = 2.52, p = 0.12.

The interaction between Face Gender and Facial Characteristics replicated Experiment 1 (Figure 12b). Participants had higher trait evaluations of feminized (Estimate = 4.63, 95% CI [4.50, 4.77]) versus masculinized (Estimate = 4.52, 95% CI [4.38, 4.65]) women's faces, b = 0.12, SE = 0.03, z = 3.93, $p < .001$. Like Experiment 1, this difference weakened among evaluations of masculinized (Estimate = 4.38, 95% CI [4.24, 4.52]), versus feminized (Estimate = 4.28, 95% CI [4.15, 4.42), men's faces b = -0.10, SE = 0.03, z = -3.32, p = .01.

Three-Way Interactions

Trait, Facial Characteristics, and Face Gender. Trait qualified the interaction between Facial Characteristics and Face Gender (Figure 13b). We characterized this interaction as we did in Experiment 1.

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

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

For women, participants similarly evaluated the dominance of feminized (Estimate = 4.42, 95% CI [4.28, 4.56]) versus masculinized (Estimate = 4.42, 95% CI [4.28, 4.56]), faces, b = -0.003, SE = 0.04, z = -0.09, $p > .99$. For men, participants evaluated masculinized (Estimate =

4.71, 95% CI [4.57, 4.85]), versus feminized (Estimate = 4.31, 95% CI [4.17, 4.45]), faces as more dominant, b = -0.40, SE = 0.04, z = -10.91, p < .001.

Trait, Face Gender, and Perceiver Gender. Trait qualified the interaction between Face Gender and Perceiver Gender (Figure 14b). We characterized this interaction as we did in Experiment 1.

Men evaluated men's faces as similarly dominant (Estimate = 4.48, 95% CI [4.33, 4.64]) and competent (Estimate = 4.40, 95% CI [4.25, 4.56]), b = -0.13, SE = 0.04, z = -3.515, p = 0.02. In contrast, they evaluated women's faces as more competent (Estimate = 4.76, 95% CI [4.60, 4.91]) than dominant (Estimate = 4.38, 95% CI [4.22, 4.53]), b = 0.54, SE = 0.04, z = 14.49, p < .001. Men evaluated men's faces as more competent (Estimate = 4.40, 95% CI [4.25, 4.56]) than warm (Estimate = 4.03, 95% CI [3.87, 4.18]), b = 0.38, SE = 0.04, z = 10.36, p <.001. They also rated women's faces as more competent (Estimate = 4.76, 95% CI [4.60, 4.91]) than warm (Estimate = 4.53, 95% CI [4.38, 4.69]), b = 0.28, SE = 0.04, z = 7.63, p <.001. Men evaluated men's faces as more dominant (Estimate = 4.48, 95% CI [4.33, 4.64]) than warm (Estimate = 4.03, 95% CI [3.87, 4.18]), b = 0.51, SE = 0.04, z = 13.88, p <.001. In contrast, they rated women's faces as more warm (Estimate = 4.53, 95% CI [4.38, 4.69]) than dominant (Estimate = 4.38, 95% CI [4.22, 4.53]), b = -0.25, SE = 0.04, z = -6.86, p <.001.

Like men, women evaluated men's faces as dominant (Estimate = 4.54, 95% CI [4.37, 4.71]) and competent (Estimate = 4.51, 95% CI [4.34, 4.67]), b = -0.03, SE = 0.03, z = -1.13, p = 0.99. Also like men, women evaluated women's faces as more competent (Estimate = 4.87, 95% CI [4.71, 5.04]) than dominant (Estimate = 4.46, 95% CI [4.30, 4.63]), b = 0.41, SE = 0.03, z = 14.56, p <.001. Like men, women evaluated men's faces as more competent (Estimate = 4.51, 95% CI [4.34, 4.67]) than warm (Estimate = 4.03, 95% CI [3.87, 4.20]), b = 0.47, SE = 0.03, z =

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

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

We again created a composite outcome rating by averaging across the three outcomes (Cronbach's $\alpha = 0.93$). We regressed these composite ratings on Job Description (-1 = analytical, 1 = interpersonal), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions as fixed effects. Perceiver Gender was included on an exploratory basis. See Table 14 for all model coefficients for the fixed effects. The model included a random effects structure that allowed intercepts to vary by participant and face identity and allowed Facial Characteristics effects to vary by participant and face identity. See Supplemental Material (C) for breakdown of each individual outcome variable.

Table 13: Effects of Face Gender, Facial Characteristics, Job Description, and Participant Gender on Evaluated Warmth (a), Competence (b), and Dominance (c) in Experiment 2.

	a. Warmth			b. Competence			c. Dominance		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*p*
(Intercept)	4.26	4.11 – 4.41	**<0.001**	4.64	4.51 – 4.76	**<0.001**	4.47	4.33 – 4.60	**<0.001**
Face Gender	0.23	0.12 – 0.34	**<0.001**	0.18	0.09 – 0.27	**<0.001**	-0.04	-0.14 – 0.05	0.375
Facial Characteristics	0.08	0.05 – 0.11	**<0.001**	0.03	0.01 – 0.06	**0.006**	-0.10	-0.14 – -0.07	**<0.001**
Job Description	0.09	-0.01 – 0.19	0.074	0.10	0.01 – 0.19	**0.037**	0.07	-0.02 – 0.16	0.112
Participant Gender	-0.02	-0.12 – 0.08	0.693	0.05	-0.04 – 0.15	0.250	0.04	-0.05 – 0.12	0.431
Face Gender * Facial Characteristics	0.03	-0.00 – 0.05	0.071	0.04	0.01 – 0.06	**0.003**	0.10	0.07 – 0.13	**<0.001**
Face Gender * Job Description	-0.01	-0.03 – 0.00	0.115	-0.01	-0.02 – 0.01	0.315	0.02	0.00 – 0.04	**0.024**
Facial Characteristics * Job Description	-0.02	-0.04 – 0.00	0.137	0.00	-0.01 – 0.02	0.646	0.02	-0.00 – 0.04	0.111

Face Gender * Participant Gender	-0.02	-0.04 – -0.01	**0.009**	0.00	-0.01 – 0.02	0.806	0.01	-0.01 – 0.02	0.454
Facial Characteristics * Participant Gender	-0.01	-0.04 – 0.01	0.144	0.00	-0.01 – 0.02	0.731	0.03	0.00 – 0.05	**0.027**
Job Description * Participant Gender	0.05	-0.05 – 0.14	0.348	0.02	-0.07 – 0.11	0.660	-0.02	-0.11 – 0.07	0.621
Face Gender * Facial Characteristics * Job Description	-0.01	-0.02 – 0.01	0.537	-0.01	-0.02 – 0.01	0.514	-0.02	-0.03 – 0.00	0.098
Face Gender * Facial Characteristics * Participant Gender	-0.01	-0.03 – 0.01	0.308	0.00	-0.01 – 0.02	0.771	-0.00	-0.02 – 0.01	0.696
Face Gender * Job Description * Participant Gender	0.01	-0.00 – 0.03	0.117	-0.00	-0.02 – 0.01	0.653	-0.01	-0.02 – 0.01	0.441
Facial Characteristics	0.01	-0.01 – 0.03	0.234	-0.00	-0.02 – 0.01	0.629	-0.01	-0.03 – 0.02	0.556

* Job Description * Participant Gender									
Face Gender * Facial Characteristics * Job Description * Participant Gender	-0.00	-0.02 – 0.01	0.662	-0.00	-0.02 – 0.02	0.978	0.01	-0.01 – 0.03	0.380

Note. Participant Gender: (-1 = Men participants, 1 = Women participants), Face Gender: (-1 = Men's faces, 1 = Women's faces), Facial Characteristics: (-1 = Masculine, 1 = Feminine), Job Description: (-1 = analytical, 1 = interpersonal)

Table 14: Effects of Face Gender, Facial Characteristics, Job Description, and Participant Gender on (a) Candidate Endorsement, (b) Likelihood of Doing Well in Position, and (c) People Liking Candidate as a Leader in Experiment 2.

	a. Candidate Endorsement			**b. Likelihood of Doing Well in Position**			**c. People Liking Candidate as a Leader**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	4.14	4.00 – 4.29	**<0.001**	4.29	4.15 – 4.42	**<0.001**	4.22	4.08 – 4.36	**<0.001**
Face Gender	0.21	0.11 – 0.32	**<0.001**	0.20	0.10 – 0.30	**<0.001**	0.18	0.08 – 0.28	**0.001**
Facial Characteristics	0.04	0.01 – 0.08	**0.010**	0.04	0.01 – 0.07	**0.009**	0.05	0.02 – 0.09	**0.001**
Job Description	0.12	0.02 – 0.22	**0.023**	0.07	-0.02 – 0.17	0.129	0.10	0.00 – 0.20	**0.041**
Participant Gender	-0.02	-0.13 – 0.08	0.656	-0.04	-0.13 – 0.06	0.456	-0.02	-0.12 – 0.07	0.622
Face Gender * Facial Characteristics	0.04	0.01 – 0.08	**0.005**	0.04	0.01 – 0.07	**0.003**	0.05	0.02 – 0.08	**<0.001**
Face Gender * Job Description	0.01	-0.01 – 0.03	0.395	0.01	-0.01 – 0.03	0.357	0.00	-0.01 – 0.02	0.680
Facial Characteristics * Job Description	0.01	-0.01 – 0.03	0.354	0.01	-0.01 – 0.03	0.489	0.00	-0.02 – 0.02	0.946
Face Gender * Participant Gender	0.01	-0.01 – 0.03	0.431	0.01	-0.01 – 0.02	0.449	-0.02	-0.04 – -0.00	**0.040**
Facial Characteristics * Participant Gender	-0.01	-0.03 – 0.01	0.316	-0.02	-0.04 – 0.01	0.141	-0.01	-0.03 – 0.02	0.497

Job Description * Participant Gender	0.01	-0.09 – 0.11	0.830	0.01	-0.08 – 0.11	0.765	-0.01	-0.11 – 0.09	0.863
Face Gender * Facial Characteristics * Job Description	-0.02	-0.04 – -0.00	**0.029**	-0.01	-0.03 – 0.00	0.109	-0.01	-0.03 – 0.01	0.343
Face Gender * Facial Characteristics * Participant Gender	0.00	-0.02 – 0.02	0.790	-0.00	-0.02 – 0.02	0.887	-0.00	-0.02 – 0.02	0.847
Face Gender * Job Description * Participant Gender	-0.00	-0.02 – 0.02	0.888	-0.01	-0.02 – 0.01	0.449	0.00	-0.02 – 0.02	0.812
Facial Characteristics * Job Description * Participant Gender	-0.01	-0.04 – 0.01	0.171	-0.01	-0.03 – 0.01	0.521	-0.00	-0.03 – 0.02	0.793
FaceGender * Facial Characteristics * Job Description * Participant Gender	-0.01	-0.02 – 0.01	0.523	-0.01	-0.02 – 0.01	0.487	-0.01	-0.03 – 0.00	0.144

Note. Participant Gender: (-1 = Men participants, 1 = Women participants), Face Gender: (-1 = Men's faces, 1 = Women's faces), Facial Characteristics: (-1 = Masculine, 1 = Feminine), Job Description: (-1 = analytical, 1 = interpersonal)

Main Effects

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

Two-Way Interaction

Replicating Experiment 1, Face Gender and Facial Characteristics interacted (Figure 15b). Participants had higher endorsements of feminized (Estimate = 4.50, 95% CI [4.33, 4.68]) versus masculinized (Estimate = 4.32, 95% CI [4.14, 4.50]), women's faces, $b = 0.18$, SE = 0.04, $z = 4.34$, $p < .001$. Participants similarly endorsed feminized (Estimate = 4.02, 95% CI [3.84, 4.19]) and masculinized (Estimate = 4.02, 95% CI [3.84, 4.20) men's faces, $b = -0.01$, SE = 0.04, $z = -0.05$, $p = 0.99$. A sensitivity analyses revealed that, given our specified model and sample size, we could detect an interactive effect size of 0.026 with 80% power and alpha = .05.

Linking Beliefs about College Students to Outcomes

We next examined whether beliefs about college students related to women versus men being more favorably evaluated. Because we wanted to explore whether *gender-biased beliefs* about college students' traits qualified Face Gender effects, we created difference scores for beliefs about warmth, competence, dominance, and responsibility by subtracting the belief about men from the belief about women. Higher scores reflect more attribution to women than men. We standardized difference scores around the mean for each trait. Participants believed that

women were warmer, $t(372) = 11.01$, $p < .002$, $d = 0.77$, more competent, $t(372) = 6.11$, $p < .001$, $d = 0.42$, and more responsible, ($t(372) = 12.26$, $p < .001$, $d = 0.97$, than men. They believed men were more dominant than women, $t(372) = -11.50$, $p < .001$, $d = 0.92$.

We regressed the composite outcome rating on Face Gender (-1 = man, 1 = woman), standardized beliefs about students' Warmth, Competence, Dominance, and Responsibility, as well as interactive effects between Face Gender and each belief. See Table 15 for coefficient information for these fixed effects. The random effects structure allowed intercepts to vary by participant and face identity.

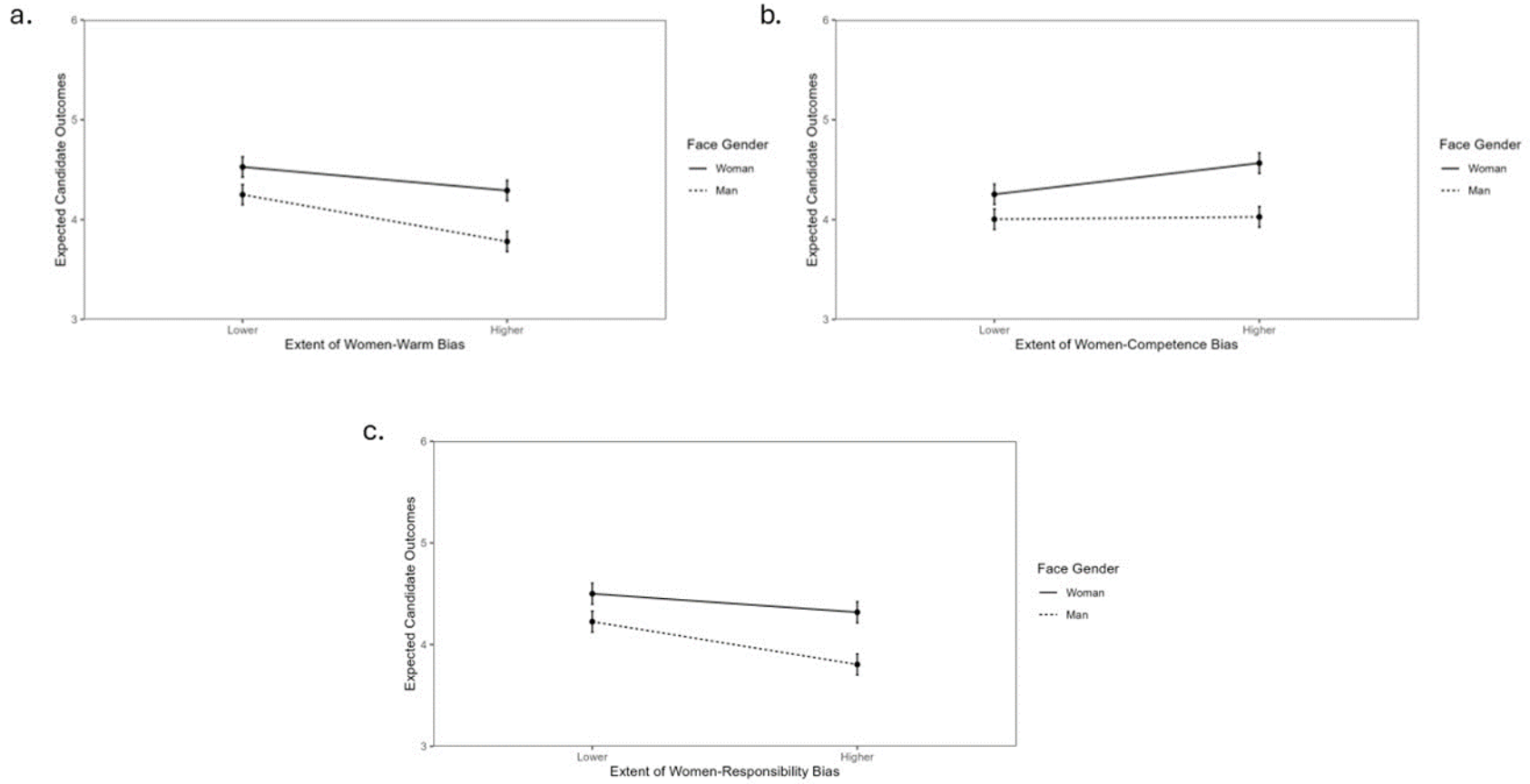
Table 15: Effects of Face Gender, warmth-related gender stereotype, competence-related gender stereotype, dominance-related gender stereotype, and responsibility-related gender stereotype on all evaluated outcomes (combined) in Experiment 2.

	Evaluated Outcomes		
Predictors	*Estimates*	*CI*	*P*
(Intercept)	4.21	4.08 – 4.35	**<0.001**
Face Gender	0.20	0.10 – 0.30	**<0.001**
Warmth-related gender stereotype	-0.18	-0.28 – -0.07	**0.001**
Competence-related gender stereotype	0.08	-0.02 – 0.19	0.128
Dominance-related gender stereotype	0.06	-0.03 – 0.16	0.203
Responsibility-related gender stereotype	-0.15	-0.26 – -0.04	**0.008**
Face Gender * Warmth-related gender stereotype	0.06	0.04 – 0.08	**<0.001**
Face Gender * Competence-related gender stereotype	0.07	0.05 – 0.09	**<0.001**

Face Gender * Dominance-related gender stereotype	-0.00	-0.02 – 0.02	0.790
Face Gender * Responsibility-related gender stereotype	0.06	0.04 – 0.08	**<0.001**

Face Gender and Warmth beliefs interacted (Figure 16a). Believing college women are warmer than men negatively related to outcomes for men, b = -0.24, SE = 0.05, z = -4.32, p <.0001. To a lesser extent, it negatively related to outcomes for women, b = -0.12, SE = 0.05, z = -2.17, p = 0.03. Face Gender and Competence beliefs interacted (Figure 16b). Believing college women are more competent than men positively related to outcomes for women, b = 0.16, SE = 0.06, z = 2.78, p = 0.005. It did not significantly relate to outcomes for men, b = 0.01, SE = 0.06, z = 0.21, p = 0.83. Face Gender and Responsibility beliefs interacted (Figure 16c). Believing college women are responsible than men negatively related to outcomes for men, b = -0.21, SE = 0.06, z = -3.65, p < .001. It did not significantly relate to outcomes for women, b = -0.09, SE = 0.06, z = -1.58, p = 0.11.

Figure 16: Face Gender by Trait-related Warmth, Competence, and Responsibility Gender Stereotypes in Experiment 2



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

Discussion

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

The extent to which people believed college-aged women versus men were more competent more positively related to outcomes for women versus men. That women were believed to be more competent than men reflects growing societal beliefs in competence equality between men and women (Eagly et al., 2020) may tip toward gender-biased beliefs in some populations. That gender-biased beliefs favorably related to evaluations of women also suggests that women do not always incur a penalty for being high in an agentic quality (i.e., competence). In fact, they may sometimes be rewarded for it. Future work may test if competency relates to favorable outcomes for women in the absence of favorably gender-biased beliefs relative to other agentic traits. Women may incur negative bias in contexts where beliefs misalign with their success. One possibility is that because beliefs about college students were assessed after the main task, the main task could have influenced these beliefs. Although we cannot be certain about this possibility, the job descriptions did not elicit any differences in gender-biased beliefs

(see Supplemental Material (C)). At the very least, our between-participants manipulation did not unduly influence beliefs. Because attitudes are malleable based on context (Blair, 2002), future work should directly test this possibility by manipulating task order. Further, because our link from beliefs to outcomes is correlational, future work can build on these findings by manipulating beliefs to determine causality.

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

Beyond examining the overall target gender effect on outcomes, Experiment 2 served as a replication of Experiment 1. The findings of Experiment 1 were largely replicated in Experiment 2 using a new and larger participant sample. Replicating Experiment 1 were trait attributions based on gender and sexually dimorphic facial features. Also replicating Experiment 1, differential evaluations only emerged for women based on sexually dimorphic facial features. Further, feminized women's faces again received the highest endorsements relative to all other combinations of facial characteristics and target gender (see Supplemental Material (C)).

General Discussion

Sexually dimorphic facial features were supported as a barrier to aspiring women leaders receiving support. Aligning with expectations from backlash research (Williams & Tiedens, 2016), women with masculinized versus feminized features received more negative evaluations

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

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

That women with masculinized versus feminized faces received more negative outcomes could be construed as nonverbal backlash. Yet, women were consistently favored over men for the student leadership role. Beliefs that college-aged women versus men were more warm, competent, and responsible positively related to these favorable outcomes for women. Notably, feminized women's faces received the most positive outcomes relative to all other combinations of target facial features and genders. Benefits for women, and especially women with feminized faces, contrasts a backlash pattern. At first blush, this possibility may seem positive for women. Yet, manipulating sexually dimorphic facial features only affected women candidates. Thus, even when women might "have it all," so to speak, they still can be penalized for possessing stereotypically masculine qualities. We suggest that sexually dimorphic facial features comprise a nonverbal barrier specific to aspiring women leaders.

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

Supporting this possibility, attractiveness confers stronger inferences of sophisticated minds (Sherman & Haidt, 2011), mental stability (Farina et al., 1977), and social value (Fink & Penton-Voak, 2002). Attractiveness also more strongly affects evaluations of women's relative

to men's mental sophistication (Alaei et al., 2022). Because feminized versus masculinized faces are evaluated as being more attractive (Perrett et al., 1998), it could be that a halo inherent to attitudes about college-aged women is accentuated among women with more feminized facial features. This halo may also reflect a preference for more stereotypically and conventionally "attractive" women reflected in their leadership prospects (Hoss et al., 2005; Hu et al., 2018; Perrett et al., 1998) versus a halo reflecting that these women are more mentally able. Future work may disentangle these possibilities. Such halos, however, are likely context-dependent. Indeed, women with feminized faces are often not considered suitable for agentic leadership roles (Banchefsky et al., 2016; Joshi et al., 2023). Future work could collect attractiveness ratings of targets and manipulating the context of the leadership role to examine a unique role for attractiveness and a potential attractiveness halo in outcomes for aspiring women leaders. If such a halo exists, it would reflect a nonverbal route contributing to disparities in leadership among women potentially separable from the backlash shown in prior work (e.g., Rudman & Glick, 1999).

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

may examine the extent to which agentic women with feminized facial features are evaluated as more agentic or less communal as they enact leadership-related behaviors.

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

Another potential limitation here is that agency- and communion-related inferences were constrained to competence, dominance, and warmth. Because agency and communion are likely not familiar terms outside of academic discourse, we used traits more familiar to a general population and that have been used in related work (Abele et al., 2016; Fiske et al., 2002, 2007; Imhoff et al., 2013; Wen et al., 2020). Future work may benefit from using different traits related to agency and communality to assess the generalizability of the current findings. For example, it

whereas warmth is an oft-tested communal trait, it can be argued that it is only a singular component of the larger concept of communion. Thus, future work may benefit from assessing different communal traits (e.g., caring) and how their inferences are differentially elicited by sexually dimorphic facial features. It could be that some communal traits are more strongly inferred by these features than others.

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

Examining contributions of sexually dimorphic facial features to outcomes for aspiring women leaders enhances our understanding of gender disparities in leadership. Having more feminized versus masculinized sexually dimorphic features benefit college-aged women regardless of whether a job is more stereotypically feminine versus masculine. Future work may

build upon the provided foundation to form a more comprehensive picture of the challenges aspiring leaders face through nonverbal cues they may not be able to control.

CHAPTER V: INTEGRATIVE DISCUSSION

The goal of the present integrative dissertation was to better understand the barriers and discrimination women face based on social and facial cues when striving for leadership positions. Here, I presented work that examined how voters may use social cues to infer perceived agency and communality of female politicians, how people expect more or less facial masculinity and femininity based on traits known or behaviors needed of that person, and how facial femininity and masculinity differentially affect women's prospects at student leadership, regardless of the agentic or communal framing of that position. Below, I discuss how my work is situated within the current literature, as well as the implications that my work has for overall better understanding the barriers women face when aspiring for leadership roles, particularly in regard to their perceived level of agency and communality. Finally, I conclude with a brief section on the applied implications of this work. Through each section of the integrative discussion, I highlight the limitations of the work reported here and discuss approaches that could be used in future research to address outstanding questions that remain.

Extending the Current Literature

With the work described in this dissertation, I contribute to the existing body of literature on the lack of women in leadership roles by focusing on how gendered trait inferences, expectations of women, and nonverbal cues may be affecting their leadership prospects. The existing literature surrounding women's challenges in attaining leadership positions is multifaceted. Prior research has suggested a connection between gendered trait inferences drawn from vignettes and faces during person perception (Oh et al., 2019; Sutherland et al., 2015; Todorov & Uleman, 2002, 2003), yet no one had directly studied this connection until the present program of research. The current dissertation enhances our understanding of how women

can effectively navigate obstacles to leadership, from political campaign framing to sexually dimorphic facial features, that influence people's perceptions and evaluations of women in regard to their perceived traits and abilities. The present studies break new ground by consistently demonstrating this connection across diverse contexts, methodologies and by evaluating real-life women (Kamala Harris), hypothetical employees, and job candidates.

To date, the findings from Chapter 3 illustrate the most direct investigation in a programmatic compilation of studies that demonstrates a nonverbal route by which people make decisions based on gender stereotypes that have wide-ranging implications for workplace behavior. We see a consistent trend emerged across six experiments of varying methodological paradigms: participants anticipated faces associated with agentic traits to exhibit masculinized features, aligning with prior research on prescriptive stereotypes and the "big two" framework (Bruckmüller & Abele, 2013a; A. E. Martin & Slepian, 2020b). These results link studies on facial trait inferences with expectations rooted in prescriptive gender stereotypes, such as recent research indicating that masculinized versus feminized facial features evoke stronger inferences of agentic traits (Walker & Wänke, 2017; Wen et al., 2020), further indicating a deep societal influence of gender on facial perception. Important to note, experiments in Chapter 3 build upon existing literature concerning facial expectations and representations rooted in gendered concepts (e.g., Degner et al., 2019; Imhoff et al., 2013) by revealing, across various scenarios, that people hold facial expectations based on traits and behaviors directly linked to sexually dimorphic facial features, but especially in regard to women.

Noteworthy was Chapter 4’s finding that aspiring female leaders with masculinized features received less favorable evaluations for leadership roles relative to female leaders with feminized features, highlighting the distinct hurdle for women posed by sexually dimorphic

facial attributes. By comparing outcome evaluations within-gender for candidates, as opposed to between-gender (Rudman & Glick, 2001; Rudman & Glick, 1999), this work expands upon existing backlash literature. Indeed, this finding can be strongly interpreted as specifically sexual dimorphism creating a unique barrier to women's leadership prospects because a within-gender paradigm allows us to hold all other aspects of the variations in facial features constant in my analyses and directly compare on only one facial modification: sexual dimorphism. The findings presented in my paper offer compelling evidence that manipulating sexually dimorphic facial features significantly affects female candidates, particularly when they possess stereotypically masculine facial features, underscoring the potential for negative evaluations when these more masculine women aspire to high-power roles.

However, some of the results in Chapter 4 were unexpected relative to prior literature on role congruity in regard to women's fit in leadership (e.g., Bosak & Sczesny, 2011), such that feminized women's faces received the highest evaluations out of all the face pairings (masculinized women, feminized men, and masculinized men). While at first counter to previous findings that men are perceived as best aligning with leadership roles and values (Badura et al., 2018; Bosak & Sczesny, 2011), I provide two possible explanations for these findings. First, feminized women's faces could have been the highest evaluated category of faces due to a "halo effect," a theory that people assume attractive people have positive personality traits and qualities (Eagly et al., 1991). Indeed, attractive faces are frequently perceived as indicating better leadership qualities because the features associated with facial attractiveness also signify good health (Thornhill & Grammer, 1999; Law Smith et al., 2006), intellectual sophistication (Sherman & Haidt, 2011), mental stability (Farina et al., 1977), and social value (Fink & Penton-Voak, 2002). Important to note, attractiveness has a stronger impact on evaluations of women's

mental sophistication compared to men's (Alaei et al., 2022), thus the halo effect may reflect a preference for conventionally attractive women in leadership roles (Hoss et al., 2005; Hu et al., 2018; Perrett et al., 1998).

A second alternative explanation for this finding of women being preferred over men for the student leadership role could stem from the perception of the leadership position's status. Given that the leadership role in the present paper was within a student context, the leadership role may not have been viewed as particularly prestigious or important compared to leadership roles in other life stages or contexts, such as STEM fields, academia, or politics (Joshi et al., 2023; Schneider et al., 2016; Schneider et al., 2021). It is plausible that people are more inclined to endorse women for leadership positions when those roles are perceived to carry lower status or significance, such as student leadership roles compared to those in the professional world. In fact, studies suggest that gender disparities in support for leadership diminish when the leadership position entails less autonomy in decision-making (Smith et al., 2007).

Conversely, even in a high-status and powerful field like politics, negative evaluations of assertive women are common in the political field (Schneider et al., 2022). Merely aspiring to political office may clash with society's expectations of women, potentially reducing support for women who put their name on the ballot (Conroy et al., 2020; Okimoto & Brescoll, 2010). In my two studies, I build upon existing research regarding how people evaluate real female candidates (e.g., Gervais & Hillard, 2011), which has primarily focused on White women (Carlin & Winfrey, 2009b; Conroy et al., 2020a; Gervais & Hillard, 2011; Ratliff et al., 2019; Rothwell et al., 2019), by investigating how perceived communality and the ideologies of the evaluators influence assessments of Kamala Harris.

Noteworthy here, this is the first investigation look at a female politician of color – Kamala Harris, who identifies as a Black and South Asian mixed-race woman – providing important insight into how emphasizing agency or communality may result in differential outcomes for women of color aspiring to politics. Perceptions of Harris's agency were heightened when people read descriptions emphasizing her agentic traits and behaviors, extending existing research on trait inferences of female politicians (Bauer & Carpinella, 2018; Conroy et al., 2020; Conroy & Green, 2020; Gervais & Hillard, 2011) by suggesting that descriptions can shape people's perceptions of female candidates of color, which is crucial given the absence of communal traits in stereotypes of Black and Asian women of color (Rosette et al., 2016). Indeed, Black women are often stereotyped as highly dominant, angry, and hard-working (Rosette et al., 2016), which differ from more communal expectations of White women (Johnson et al., 2012b), putting women of color up against unique challenges due to their intersecting racial and gender identities (Gershon & Lavariega Monforti, 2021).

Due to this clash of stereotypes, some research suggests that perceptions of communality may not strongly influence evaluations of agentic Black women (Livingston et al., 2012), while other work has shown these differing stereotypes of women of color to contribute to poorer leadership assessments for Black women compared to White women (Motro et al., 2021). However, based on my work in this dissertation, I posit that communality may actually be crucial for women of color to mitigate penalties associated with their agency, as seen with perceived communality having a strong positive relation with evaluations of Kamala Harris. The significant contribution of perceived communality to evaluations of Harris across both of my studies supports this notion that emphasizing communality may be particularly vital for women of color

striving for high-status leadership positions in order to combat agentic stereotypes of women of color and receive positive evaluations from voters.

According to an identity-based model of political ideology (Van Bavel & Pereira, 2018), the effects of communality on evaluations of Harris should vary depending on whether perceivers lean more conservative or liberal. Indeed, both liberals and conservatives typically make negative judgments about people whose values don't align with their own (Morgan et al., 2010). In prior work, both perceived communality and a more liberal ideological stance were found to positively influence the likelihood of voting for Hillary Clinton in the 2008 election, even after accounting for sexism (Gervais & Hillard, 2011). Therefore, we explored the interactive relationship between these factors to understand how communality impacts evaluations of other real-life female candidates, and found differing results. While I found liberals to be more in favor of Harris overall, liberals evaluated Harris more positively when they also perceived her as low in communality – but only in the context when participants had just read a description emphasizing Harris's communality.

This finding shows that positive perceptions of communality had a weaker impact on evaluations of Kamala Harris among individuals with more liberal versus conservative ideologies, aligning with recent work demonstrating that candidates emphasizing traits traditionally associated with femininity do not necessarily pose a barrier for women seeking political office to be elected (Blais & Sevi, 2024). Indeed, we extend this in Chapter 2 to show how female candidates can emphasize stereotypically feminine traits in certain contexts (i.e., the communal context) to their advantage for certain audiences (towards liberal versus conservative voters).

Applied Practical Implications

Past work asserts that female candidates have a tendency to emphasize traditional feminine stereotypes during their campaigns (Schneider 2014), potentially leading to adverse outcomes during elections. Other work asserts that female candidates might find it advantageous to emphasize masculine attributes (Sapiro et al. 2011), since traits traditionally linked with femininity, like warmth and compassion, may not be deemed appropriate for a leader in politics. However, findings from Chapter 2, that people who perceived Harris as more communal had more positive evaluations of her (despite her being perceived as more agentic overall), align with a recent investigation conducted by Bauer and Santia (2022) that female candidates should strategically manage the portrayal of both feminine and masculine characteristics during their campaigns to mitigate potential likability backlash. Indeed, as previously mentioned, positive communality effects on evaluations of Harris were weaker for more liberal, compared to more conservative, participants after reading a communal description of her. This evidence of varying levels of support based on communal or agentic emphasis have practical implications that female candidates of color perhaps cannot rely solely on support from those who share their political beliefs, as gendered trait inferences from campaigns advertisements (such as emphasizing family values versus career history) may alter evaluations against them. Additionally, these results demonstrate a scenario in which the framing of a female politician's political campaigns (such as emphasizing family history and values versus career accomplishments) can influence voters, despite the candidate aligning with their political party.

Another implication is based on findings from Chapter 3, which demonstrates how gendered expectations of people may prevent certain individuals from opportunities in the workplace and beyond, such as expecting masculine people to be best for solving technical

problems, but feminine people to be best for solving interpersonal problems (as specifically demonstrated in Experiment 3b). Aligning with work that people believe men are best suited for more stereotypically masculine jobs and women are best suited for stereotypically feminine jobs (Glick et al., 1995), these expectations can keep women from opportunities and promotions at work based on gendered assumptions of their skills, when these assumptions could in fact be incorrect. Intervention efforts aimed at reshaping societal perceptions of gender should include a focus on facial cues, as these nonverbal cues often trigger gender stereotypical inferences. Relevant here, recent work had adult participants undergo counterstereotype training to reduce their reliance on facial appearance in social judgments (Hong et al., 2024). The training diminished the predictive power of trustworthiness and prison sentencing judgments on real-world inmate outcomes, demonstrating the training's potential effectiveness to realistic decision-making contexts with important consequences. In tandem with recent relevant work that found trustworthiness inferences based on face perception were updated only when the updated information was extreme and reliably believable (Shen & Ferguson, 2021), these findings give hope to the effectiveness of actual interventions being implanted to combat negative social judgements based on face perception. Yet, no work has examined interventions like this in regard to combatting gender stereotypes. Therefore, similar interventions targeting gendered trait inferences during face perception could play a pivotal role in challenging and altering entrenched gender biases potentially obstructing women's path to leadership positions.

Moreover, an early-childhood intervention concerning the conceptualizations of gender groups could be particularly crucial for young children as they are beginning to form their understanding of gendered groups and stereotypes. In Chapter 4, I present evidence of negative leadership outcomes for women who exhibit more masculine versus more feminine facial

features. A practical implication from this body of work is how deeply engrained gendered appearance expectations are of women in our society, prompting a need for interventions targeting gender biases from an early age. Beginning interventions at a young age in order to unlearn and better understand how people perpetuate gender stereotypes, as well as how people make gendered stereotypical inferences about others from nonverbal cues might serve as an initial stride towards reshaping society's conceptions of gender stereotypes. Previous research indicates that children begin to identify gender groups as young as 2-years-old, and developing a relatively firm grasp on gender identities (such as being a feminist, for example) begin to form before adolescence (see e.g., Phinney & Tarver, 1988). Likewise, understanding gendered stereotypical expectations, both about themselves (Banchefsky et al., 2016) and others (Boseovski & Yuly-Youngblood, 2016) based on their perceived gender category can start to develop around early teenage years. Thus, while implementing a comprehensive lesson plan aimed at debunking gender stereotypes in the American public school system may pose challenges, it seems crucial to start earlier in childhood as opposed to later – as stereotypical expectations of gender groups have already been formed by the time people reach adulthood. Considering recent research, such as the prior work mentioned above and the present dissertation papers, when developing interventions could be a valuable effort in reducing biases in future generations' evaluations of women.

Future Work

Our findings suggest a number of possible avenues for future investigation. While this dissertation has emphasized nonverbal barriers beyond women's control in their pursuit of leadership positions, I overlook the discussion of nonverbal cues that women can intentionally manage. Makeup is one of the foremost and frequently adopted methods for altering one's facial

appearance and is predominately used by women (Waldersee, 2019). Prior work has consistently demonstrated the impact of cosmetics on perceived level of facial attractiveness for women (Batres et al., 2018; Nash et al., 2006; Mulhern et al., 2003). Makeup serves to accentuate three visual features associated with youth and health: skin uniformity, facial contrast, and facial feature size (Batres et al., 2019; Jones et al., 2018; Stephen & McKeegan, 2010). Moreover, makeup influences social perception by modifying visual indicators of physical well-being (e.g., estrogen levels, sexual arousal, cardiac and respiratory health; Russell et al., 2016; Stephen & McKeegan, 2010). The association between makeup and youthfulness, femininity, physical well-being, and attractiveness has been extensively documented (Davis & Arnocky, 2020).

Nonetheless, the impact of facial makeup on women's prospects in leadership roles remains relatively unexplored. Recent work supports that faces with light makeup are more easily recognized versus faces with heavy makeup (Tagai et al., 2016) and women's faces with light makeup are evaluated as more attractive and more fluently processed (Tagai et al., 2017). Important to note, the fluency in which a face is processed impacts the likability of that person (Lanska et al., 2014; Lick & Johnson, 2013, 2015; Olszanowski et al., 2018). These recent results suggest that the processing fluency of faces with light makeup is an important reason why women's faces with light makeup are preferred to women's faces with heavy makeup and women's faces with no makeup. Indeed, women who wear heavier makeup are associated with perceived higher sexual activity and attractiveness, but is also associated with the perception that these women have lower mental capacity and moral status (Kellie et al., 2021).

Therefore, building on the current dissertation's body of work, women's level of makeup use could drastically affect how she is evaluated for a leadership position. Indeed, if a woman wears heavy makeup, she may be evaluated as unfit for a leadership role due to the association

that women with heavy makeup have lower mental capacity (Kellie et al., 2021). Moreover, makeup use may differ in impact for women's leadership prospects depending on the field and career path, such as a job in medicine versus politics. Relevant to the present dissertation, perhaps female politicians need to consider makeup use's influence on perceptions of leadership ability during political campaigns.

Despite prior work that heavy makeup consistently reduces perceived leadership ability in women, regardless of race (James et al., 2018), makeup use could have differential outcomes for Black women versus for White women in a political setting. To address this gap in research and increase external validity within the field, future studies should explore the impact of makeup on real-world women from diverse racial backgrounds striving for leadership roles, particularly in the context of political campaigns. For example, using a paradigm similar to the project in Chapter 4, a future study could fill the aforementioned gap by having participants evaluate political candidates (using racially diverse face images of women) with no makeup, light makeup, and heavy makeup to see who is best suited and most likely to be elected for a local political position. Based on previous findings on Black women's hairstyles (Karl et al., 2022), I would expect in a conservative environment like the field of politics, Black women would actually incur more negative evaluations for no makeup use compared to White women, as adjusting their appearance by using makeup to mirror White, Eurocentric beauty standards would actually be viewed as more "professional" for Black women compared to White women, for whom more makeup use may be perceived as "unprofessional."

Consequently, makeup works by adjusting biologically connected elements of attractiveness, such as the youthful appearance of one's face (Jones et al., 2018), offering older women a more youthful look and younger women a more mature one. Consequently, the impact

of makeup on women's leadership prospects may vary depending on their age and life stage. Age undeniably plays a role in how leadership readiness is perceived. For instance, research suggests that older, more experienced leaders are preferred in stable times, while younger, more exploratory leaders are favored in times of change (Spisak et al., 2014). However, the relationship between facial aging and experience may not translate the same way for women. Indeed, younger men are overrepresented in leadership roles compared to older men and women in general (Tresh et al., 2019).

According to the stereotype content model (Fiske et al., 1999, 2002), age stereotypes are aligned with the warmth-competence dimensions, portraying older individuals as warmer but less competent. These stereotypes persist in the workplace, where older workers are often perceived as having lower performance, technical skills, and productivity compared to their younger counterparts (Broadbridge, 2001; Cuddy and Fiske, 2002; Posthuma and Campion, 2009). While warmth stereotypes generally view older workers positively as loyal and socially adept (Warr and Pennington, 1993), leadership selection tends to favor competence-related traits, placing younger candidates at an advantage even when older candidates possess equivalent qualifications (Perry et al., 2017). The intersectionality of gender and age stereotypes may trigger in-group bias favoring younger, higher-status workers over their older counterparts, as competence-related age stereotypes confer a leadership advantage to younger individuals (Finkelstein et al., 1995; Gordon and Arvey, 2004). Given that facial appearance heavily influences age perceptions (Foos & Clark, 2011), women might pay particular attention to their facial features (through makeup or cosmetic procedures) to counteract these negative assessments based on aging. However, the exploration of this concept concerning subsequent inferences of gendered traits regarding women's leadership potential remains relatively unexplored (Scheuer & Loughlin, 2020), even

though there is strong evidence from the current dissertation that demonstrates the strength of influence women's faces have on their judgement of character and leadership prospects. Using a similar paradigm to Chapter 3's Experiments 3a and 3b, a future study could investigate (using face images of young and old men and women) who is perceived as best suited for certain gendered workplace tasks or leadership requirements. Based on the aforementioned prior work, I would expect older women to be *least* expected to fulfill masculine workplace tasks and leadership requirements based on the perception that older individuals are less competent compared to younger individuals. Follow-up studies could then investigate the interplay of makeup use on older adults and if this combats high communal/low agentic assumptions of older women, which would demonstrate if women can use makeup as an effective tool to challenge gendered trait expectations in the workplace at various life stages.

Another avenue to future work could be to examine how familiarity and categorization fluency could influence the trend of negative assessments of masculinized female faces that we see in the current dissertation and previous aforementioned work. Perhaps if counter-stereotypical (i.e., masculine) women are encountered less frequently their stereotypical (i.e. feminized) counterparts, then they may be less fluently processed and subsequently more negatively evaluated (Lick & Johnson, 2013). Indeed, the processing fluency model posits that increased exposure to an item or person will enhance the ease with which people recognize and process the item or person (Seamon et al., 1995). Essentially, increased fluency leads people to have more positive attitudes toward an item or person, while less exposure leads people to have more negative attitudes toward an item or person. Multiple studies have shown that fluent processing leads perceivers to evaluate items like art more favorably (Belke et al., 2010), consider instructions simpler to follow (Song & Schwarz, 2008), rate food as less risky (Song &

Schwarz, 2009), judge moral violations as less offensive (Laham et al., 2009), and deem currencies to be more valuable (Alter & Oppenheimer, 2008). This phenomenon of people preferring prototypical stimuli (Winkielman et al., 2006) is known as the *beauty-in-averageness* effect.

This theory could indeed extend to feminine women and masculine men, both of which are prototypical and thus easy to process (e.g., Sofer et al., 2015). Processing ease informs evaluative judgments of diverse stimuli, such that fluency tends to elicit positive evaluations whereas disfluency tends to elicit negative evaluations. Indeed, counter-stereotypical faces may be less familiar due to inherent facial structural features or because they are less represented or portrayed biasedly in our cultural climate (e.g., potentially due to the underrepresentation of female leaders in general or in the media: Cracknell, 2013). Investigating this question of processing ease for more feminine or masculine women's faces could be conducted through an Implicit Association Task (IAT) paradigm (Greenwald et al., 2002), with participants tasked with sorting more feminized and masculinized faces with categories (e.g., "men" and "women", "agentic" and "communal", etc.). Reaction times could inform if processing ease in categorizing masculine women as "women" and attributing them stereotypically feminine traits (i.e. "communal") contributes to the negative evaluations of more masculine women.

Another important consideration for future work is understanding the nuanced difference between subtle and blatant support for women candidates whilst they strive for leadership roles. Subtle forms of support, such as begrudgingly voting for a candidate while explicitly verbalizing one's anger or distaste for the candidate can subtly undermine women who do reach leadership positions by reinforcing stereotypes that women are not competent enough for leadership or diminishing women leaders' likability. On the other hand, blatant support, while seemingly

advantageous, may also trigger backlash effects or accusations of preferential treatment towards women in an effort for equality, further complicating the evaluation process. Therefore, discerning between these forms of support and the subsequent impact on candidate's performance in their leadership position would be beneficial for future work to investigate.

There is currently no work investigating how subtle versus blatant support during candidacy affects women leaders' actual performance outcomes once elected or hired. An informative future direction, for example, would be longitudinal studies observing the career trajectory and achievements of both men and women in leadership roles, who either experienced either overtly positive support or subtle, adverse support during their hiring phases. Future work like this could provide valuable insights into the enduring impact of blatant versus subtle support on their quality of leadership and effectiveness during their time in the role. These studies could delve into whether these leaders received promotions or salary adjustments (similar to the methodological paradigm of Chapter 3's Experiment 3a and 3b), illuminating the differential outcomes stemming from distinct types of support during their initial hiring stages. Since women often encounter unique challenges and biases in leadership contexts, I believe the exploration of how subtle versus blatant support for a female candidate leading up to her leadership hiring or election to be a fruitful and important consideration for future work and practical implications.

Strengths and Limitations of the Current Research

Due to the timing of my doctoral research beginning in 2020 during the height of the COVID-19 pandemic, all three projects in this dissertation were conducted online through a CloudResearch extension on MTurk. Specifically for gender stereotyping research, sampling from a wider population of the United States through online studies was actually advantageous in that the results are more generalizable to the general American population (Hauser et al., 2023;

Keith & McKay, 2024). Since UNCG is a predominately liberal university and the student participant SONA pool is predominately women, using online samples for the present work was beneficial to investigate a more representative sample of people with varying personal and political views.

To address concerns regarding data quality (e.g, participants are bots, not real people; Webb & Tangney, 2022), we specifically use CloudResearch for our online data sampling. Indeed, a substantial amount of research has begun exploring the reliability of MTurk samples, including direct comparisons between MTurk samples and samples from alternative sources, such as Prolific and college student samples (e.g., Berinsky et al., 2012; Keith et al., 2022) and various assessments of data-quality gathered on MTurk (e.g., Peer et al., 2022; Walter et al., 2019). Certain choices can be made during the research design, data-collection, and data-cleaning process can influence generalizability and data quality (Keith et al., 2017), such as using several attention check questions throughout the study and excluding any participants that fail even one attention check. Thus, in the present studies reported in my dissertation, I enacted best practices supported by recent research (e.g., Keith & McKay, 2024) to ensure the best possible data quality by utilizing attention checks and tools provided by CloudResearch. Indeed, CloudResearch uses anti-fraud technology and algorithms to detect suspicious activity and patterns indicative of fraudulent behavior (Hauser et al., 2023). Across multiple metrics, CloudResearch participants have been found to have superior performance to regular MTurk participants, in that they were more accurate in identifying image content, answered more reading comprehension questions correctly, showed greater consistency in responding to reverse-coded items, passed a higher number of attention checks, reported less cheating, and were less likely to leave the survey window on easily searchable questions (Hauser et al., 2023). Therefore,

we sampled our participants through CloudResearch for the present dissertation's experiments to ensure that our projects were abiding by best research practices.

Regardless, a limitation of online sampling that still remains is that our experiments' samples are predominantly White and educated. To ensure more racially and educational diversity, we could utilize CloudResearch's tools that enables researchers to specifically sample certain demographic populations (e.g., only survey participants who identify as Black). The importance of diversity in research participant populations has been highly and widely discussed in regard to psychology research (e.g., Roberts, 2024; Roberts et al., 2020). Indeed, people are socialized differently based on their race (Pauker et al., 2018), which can lead to variations in traits and behaviors based on race (e.g., McGorray et al., 2023), which are important for researchers to consider across all social psychology areas. Practically, the lack of racial diversity in psychology will leave the field unprepared for an increasingly diverse society, leaving research findings to be ungeneralizable to the increasingly diverse American population (Roberts et al., 2020).

Down a similar vein, another disadvantage of the present battery of work is the limited use of non-White faces. Due to the notion that race is gendered (Johnson et al., 2012b), which posits that each racial group has gendered associations (e.g., Black individuals associated with masculinity, East Asian individuals associated with femininity, etc.), it is important to use non-White faces in similar future work to better understand how the intersectionality of gendered and racial trait inferences differentially influence leaders who are women of color. Indeed, prior work investigating perceptions and evaluations of Barack Obama during the 2008 election found that endorsing pictures of Barack Obama with lighter skin tones as best representing him related to more intentions to vote for him (Caruso et al., 2009), plausibly because of the positive

stereotypes associated with lighter skin (Maddox & Gray, 2002). Relevant to women's leadership, prior work using a woman of color as face image stimuli helped identify factors affecting evaluations of women of color seeking political leadership (Cassidy & Liebenow, 2021a). By examining how people mentally represent Kamala Harris's face, which has been found to be an important factor affecting candidate evaluations (e.g. Young et al., 2014), results showed positive evaluations of Harris related to participants more positively endorsing feminine representations of Harris's face as best representing her. Perceiving Harris as highly feminine may have been significantly beneficial to Harris's positive evaluations as a Black woman, as more feminized features elicit communal trait inferences (e.g., Wen et al., 2020), and therefore may counteract the agentic-Black stereotype (Johnson et al., 2012b) which would otherwise hinder a Black woman's leadership prospects. Work like this supports the need for more future work to use racially diverse faces in order to discern the differing benefits of feminine-communal and masculine-agentic emphasis for leadership candidates of different racial groups.

Moreover, facial cues can be enduring and consequential in nature (e.g., Ambady & Rosenthal, 1992), with quick judgments of character from face perception resulting in important or permanent decisions. Prior work suggests that snap judgments about individuals through "thin-slicing," the act of forming impressions of others from short interactions or little information, can be long-lasting, with varying degrees in accuracy (Ambady & Rosenthal, 1993, 1997). Impressions from faces have been found to even often overwrite other incoming information that may be learned later on about that person (Jaeger et al., 2019). Despite the strong evidence of people forming impressions of others based on nonverbal cues, there is an open question in regard to if the resulting *backlash* from nonverbal cues is long-lasting, as well. This dissertation demonstrates this concept through all three papers. In Chapter 2, participants

made judgments about Kamala Harris based on brief vignettes emphasizing either her agentic or communal nature. Chapter 3 involved participants making quick assessments of facial expectations based on traits, as well as judgments about awards/tasks based on nonverbal cues from various White adult faces. In Chapter 4, participants determined who to elect as student leader after brief exposure to job descriptions and face images. These studies illustrate how quick judgments of character can result in actions with lasting consequences, such as missing out on awards or promotions due to inaccurate character assessments. Previous research supports these findings, showing that facial judgments influence important decisions with long-lasting consequences, such as appointing a CEO of a company (Gomulya et al., 2016) or sentencing a criminal to prison (Eberhardt et al., 2006; Wilson & Rule, 2015). Drawing from this collective body of research, both in the preexisting literature and in this dissertation, I assert that judgments of others and the resulting backlash and impact on their lives and careers would endure over time.

In fact, a methodological strong suit in the present dissertation is my utilization of varied methodologies and paradigms used (evaluations of face images, vignettes, and resumes) in order to better understand factors contributing to women's underrepresentation in leadership. By using both vignettes and images of a candidate (Chapter 2), face image stimuli (Chapters 3), and both job description and face image stimuli (Chapter 4) in the present body of work, instead of solely using one method, we were able to capture nuances in person perception (e.g., trait inferences from sexually dimorphic facial features) that may be otherwise difficult for people to individually detect on their own. Indeed, explicit assessments, such as participants rating face image stimuli varying in masculinity and femininity across a spectrum of traits, has been shown to be a strong and useful methodology in a variety of previous work (Walker & Wänke, 2017),

and also allows us to directly test how sexual dimorphism in faces can affect people's assumed and inferred traits about others based on their facial features. In order to make this battery of work stronger, using implicit measures, such as employing the traditional seven-step Implicit Association Task (IAT), would help in revealing participants' subconscious associations of masculine and feminine faces with agentic and communal traits, respectively (Wen et al., 2020), better indicating the ingrained nature of these gender-related associations with sexually differentiated facial features.

Furthermore, for some readers, a potential limitation of the current work may be a concern about demand characteristics. Demand characteristics refer to cues or aspects of a study that lead participants to infer the purpose or hypothesis of the research and subsequently adjust their behavior or responses to align with their perceptions of what is expected of them (Coles et al., 2023; Coles & Frank, 2023). These cues can be explicit, such as verbal instructions or the design of the study, or implicit, such as the context of the study. When participants become aware of the research objectives or hypotheses, they may consciously or unconsciously alter their behavior or responses to conform to what they believe the researcher wants or expects to find.

A hypothetical example of demand characteristics would be a study where participants evaluate both men's and women's faces simultaneously, and participants might deduce that the study focuses on gender-related evaluations and stereotypes. Thus, these hypothetical participants may alter their responses to either align with what they think the researcher's experimental hypothesis is, or to not appear sexist. Demand characteristics are an on-going concern in the social psychology field in general, with some meta-analyses seeking to investigate how demand characteristics may affect highly-reputable theories, such as the facial feedback hypothesis (Coles et al., 2023). They found that, despite some evidence for demand

characteristics, their results indicated that facial feedback effects were not solely driven by demand or placebo effects. While this is positive evidence for this theory, no similar work has yet to be done in regard to face perception or gendered trait inferences research.

In my present work, I purposefully begin my studies with a vague, yet straight-forward instruction as to what participants can generally expect to do in order to minimize the effects of potential demand characteristics. For all three of my papers, instructions were purposefully short and ambiguous, in order to only give participants enough information to do the experiment correctly, but not enough information to give away the study's purpose. For example, these were the instructions in the first paper's first experiment: "In this experiment, we are interested in how people evaluate others who are familiar to them. You will be making evaluations about Kamala Harris. First, you will read some information about Kamala Harris. Then, you will make some evaluations about her. There are no right or wrong responses." From these instructions, an understanding that we are trying to detect gendered trait perceptions of Kamala Harris is difficult to tell. While participants indeed may adjust their answers to avoid appearing sexist against a female candidate, there is no evidence that this occurred in the present work. Nevertheless, it is important to consider demand characteristics when designing and executing future work, as avoiding demand characteristics can enhance the validity and reliability of researchers' findings. For example, implementing a wide-range of measures, such as both explicit and implicit measures used in Chapter 3, can help diminish concern for demand characteristics. Indeed, the implicit nature of the reverse correlation task (Experiment 4), which found consistent results with the previous five studies in the paper, demonstrated that demand characteristics were most likely not driving the results found in Chapter 3. Future work could also include demand characteristics checks (similar to attention checks) in the form of a short answer question at the conclusion of

studies to gauge participants' perceptions of the study's purpose. However, participants' knowledge of a study's purpose does not automatically mean they will respond to align with researchers' hypotheses. Indeed, alternatively, participants may alter their responses in order to not appear as biased or prejudiced toward a group, which is a common concern in stereotype and prejudice work utilizing self-report measures (e.g., Monteith & Voils, 1998). Despite this, if demand characteristics are a heightened concern, researchers could consider excluding participants whose assumptions correctly align with the study's purpose in order to reduce the probability of effects of demand characteristics.

Conclusion

The three empirical papers reported here had the overarching goal of understanding how gendered trait and behavior inferences from public framing and facial cues can affect leadership outcomes for women. Across the studies, there was evidence that women's perceived high agency or high facial masculinity had a noticeable negative influence on their leadership prospects or candidate outcomes. Additionally, the present work demonstrates how the challenges women face when striving for leadership often stem from societal expectations regarding gendered traits and behavior, which can create conflicting standards for how women should behave as leaders. By uncovering that factors, such as agentic or communal political framing and sexually dimorphic facial features, influence people's impressions and evaluations of women, the current battery of research contributes to our understanding of how women navigate these obstacles.

While this dissertation sheds light on how gendered trait inferences and expectations of women impact perceptions of women striving for leadership, identifying the key factors that influence leadership hiring and voting decisions. As mentioned previously, future work should

incorporate more diverse facial stimuli (e.g., other race faces) and investigate how prominent appearance modifications (e.g., makeup) and natural appearance changes over time (i.e., aging) can contribute to trait inferences about these women and women's leadership evaluations and overall outcomes. By continuing to understand the roles of sexually dimorphic facial features, such as in regard to appearance alteration and aging, we can more comprehensively understand how women are being evaluated based on social framing and facial cues and, hopefully, apply practical interventions early on in future generations to prevent the prevalence of backlash from gender stereotypes. The present and future work can hopefully inform intervention strategies that aim to mitigate discrimination against women throughout their careers as they strive for positions of authority. Ultimately, having a more comprehensive understanding of gender biases in leadership election and hiring processes empowers us to pave the way for more equitable opportunities for women. This proactive approach is crucial for narrowing the gender gap in leadership roles within our society, fostering a more inclusive and diverse landscape for future generations.

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

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*(7), 1063–1065. https://doi.org/10.1002/ejsp.574

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—communion with warmth and morality. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Alaei, R., Deska, J. C., Hugenberg, K., & Rule, N. O. (2022). People attribute humanness to men and women differently based on their facial appearance. *Journal of Personality and Social Psychology*, *123*, 400–422. https://doi.org/10.1037/pspi0000364

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

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431–441. https://doi.org/10.1037/0022-3514.64.3.431

Ambady, N., & Rosenthal, R. (1997). Judging Social Behavior Using “Thin Slices.” *CHANCE*, *10*(4), 12–51. https://doi.org/10.1080/09332480.1997.10542056

Archer, J. (1996). Sex differences in social behavior. Are the social role and evolutionary explanations compatible? *The American Psychologist*, *51*(9), 909–917. https://doi.org/10.1037//0003-066x.51.9.909

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

Badura, K. L., Grijalva, E., Newman, D. A., Yan, T. T., & Jeon, G. (2018). Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology*, *71*(3), 335–367. https://doi.org/10.1111/peps.12266

Balliet, D., Tybur, J. M., Wu, J., Antonellis, C., & Van Lange, P. A. M. (2018). Political ideology, trust, and cooperation: In-group favoritism among republicans and democrats during a US national election. *Journal of Conflict Resolution*, *62*(4), 797–818. https://doi.org/10.1177/0022002716658694

Banchefsky, S., Westfall, J., Park, B., & Judd, C. M. (2016a). But you don’t look like a scientist!: Women scientists with feminine appearance are deemed less likely to be scientists. *Sex Roles*, *75*(3–4), 95–109. https://doi.org/10.1007/s11199-016-0586-1

Banchefsky, S., Westfall, J., Park, B., & Judd, C. M. (2016b). But you don't look like a scientist!: Women scientists with feminine appearance are deemed less likely to be scientists. *Sex Roles*, *75*(3–4), 95–109. https://doi.org/10.1007/s11199-016-0586-1

Baskaran, T., & Hessami, Z. (2018). Does the election of a female leader clear the way for more women in politics? *American Economic Journal: Economic Policy*, *10*(3), 95–121. https://doi.org/10.1257/pol.20170045

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, N. M. (2017). The Effects of Counterstereotypic Gender Strategies on Candidate Evaluations. *Political Psychology*, *38*(2), 279–295. https://doi.org/10.1111/pops.12351

Bauer, N. M., & Carpinella, C. (2018). Visual Information and Candidate Evaluations: The Influence of Feminine and Masculine Images on Support for Female Candidates. *Political Research Quarterly*, *71*(2), 395–407. https://doi.org/10.1177/1065912917738579

Belanger, A. L., Joshi, M. P., Fuesting, M. A., Weisgram, E. S., Claypool, H. M., & Diekman, A. B. (2020). Putting Belonging in Context: Communal Affordances Signal Belonging in STEM. *Personality and Social Psychology Bulletin*, *46*(8), 1186–1204. https://doi.org/10.1177/0146167219897181

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.

Betz, D., & Sekaquaptewa, D. (2012). My fair physicist? Feminine math and science role models demotivate young girls. *Social Psychological and Personality Science*. https://journals.sagepub.com/doi/abs/10.1177/1948550612440735

Blair, I. V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Blais, A., & Sevi, S. (2024). Are feminine traits a liability in elections? *Acta Politica*. https://doi.org/10.1057/s41269-024-00338-6

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Bligh, M. C., Schlehofer, M. M., Casad, B. J., & Gaffney, A. M. (2012a). Competent enough, but would you vote for her? Gender stereotypes and media influences on perceptions of women politicians. *Journal of Applied Social Psychology*, *42*(3), 560–597. https://doi.org/10.1111/j.1559-1816.2011.00781.x

Bligh, M. C., Schlehofer, M. M., Casad, B. J., & Gaffney, A. M. (2012b). Competent enough, but would you vote for her? Gender stereotypes and media influences on perceptions of women politicians: Women politicians. *Journal of Applied Social Psychology*, *42*(3), 560–597. https://doi.org/10.1111/j.1559-1816.2011.00781.x

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

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The Ideological-Conflict Hypothesis: Intolerance Among Both Liberals and Conservatives. *Current Directions in Psychological Science*, *23*(1), 27–34. https://doi.org/10.1177/0963721413510932

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

Broockman, D. E. (2014). Do female politicians empower women to vote or run for office? A regression discontinuity approach. *Electoral Studies*, *34*, 190–204. https://doi.org/10.1016/j.electstud.2013.10.002

Brown, E. R., Thoman, D. B., Smith, J. L., & Diekman, 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. (2013a). 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

Bruckmüller, S., & Abele, A. E. (2013b). 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

Carbon, C.-C. (2011). The first 100 milliseconds of a face: On the microgenesis of early face processing. *Perceptual and Motor Skills*, *113*(3), 859–874. https://doi.org/10.2466/07.17.22.PMS.113.6.859-874

Carlin, D. B., & Winfrey, K. L. (2009a). Have You Come a Long Way, Baby? Hillary Clinton, Sarah Palin, and Sexism in 2008 Campaign Coverage. *Communication Studies*, *60*(4), 326–343. https://doi.org/10.1080/10510970903109904

Carlin, D. B., & Winfrey, K. L. (2009b). Have you come a long way, baby? Hillary Clinton, Sarah Palin, and sexism in 2008 campaign coverage. *Communication Studies*, *60*(4), 326–343. https://doi.org/10.1080/10510970903109904

Carpinella, C. M., & Johnson, K. L. (2016). Visual political communication: The impact of facial cues from social constituencies to personal pocketbooks. *Social and Personality Psychology Compass*, *10*(5), 281–297. https://doi.org/10.1111/spc3.12249

Cassese, E., Conroy, M., Mehta, D., & Nestor, F. (2022). Media Coverage of Female Candidates' Traits in the 2020 Democratic Presidential Primary. *Journal of Women, Politics & Policy*, *43*(1), 42–63. https://doi.org/10.1080/1554477X.2022.2004846

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., & Krendl, A. C. (2019). A crisis of competence: Benevolent sexism affects evaluations of women's competence. *Sex Roles*, *81*(7–8), 505–520. https://doi.org/10.1007/s11199-019-1011-3

Cassidy, B. S., & Liebenow, H. A. (2021a). 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

Cassidy, B. S., & Liebenow, H. A. (2021b). 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., & Mull, M. S. (2006). Conservative Ideology and Ambivalent Sexism. *Psychology of Women Quarterly*, *30*(2), 223–230. https://doi.org/10.1111/j.1471-6402.2006.00284.x

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

Chua, K.-W., & Freeman, J. B. (2021). Facial Stereotype Bias Is Mitigated by Training. *Social Psychological and Personality Science*, *12*(7), 1335–1344. https://doi.org/10.1177/1948550620972550

Clayton, K., Crabtree, C., & Horiuchi, Y. (2022). Do Identity Frames Impact Support for Multiracial Candidates? The Case of Kamala Harris. *Journal of Experimental Political Science*, 1–12. https://doi.org/10.1017/XPS.2021.33

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

Coles, N. A., & Frank, M. C. (2023). *A quantitative review of demand characteristics and their underlying mechanisms*. https://doi.org/10.31234/osf.io/uw85a

Coles, N. A., Gaertner, L., Frohlich, B., Larsen, J. T., & Basnight-Brown, D. M. (2023). Fact or artifact? Demand characteristics and participants' beliefs can moderate, but do not fully account for, the effects of facial feedback on emotional experience. *Journal of Personality and Social Psychology*, *124*(2), 287–310. https://doi.org/10.1037/pspa0000316

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

Conroy, M., & Green, J. (2020). It Takes a Motive: Communal and Agentic Articulated Interest and Candidate Emergence. *Political Research Quarterly*, *73*(4), 942–956. https://doi.org/10.1177/1065912920933668

Conroy, M., Joesten Martin, D., & Nalder, K. L. (2020a). Gender, sex, and the role of stereotypes in evaluations of Hillary Clinton and the 2016 presidential candidates. *Journal of Women, Politics & Policy*, *41*(2), 194–218. https://doi.org/10.1080/1554477X.2020.1731280

Conroy, M., Joesten Martin, D., & Nalder, K. L. (2020b). Gender, sex, and the role of stereotypes in evaluations of Hillary Clinton and the 2016 presidential candidates. *Journal of Women, Politics & Policy*, *41*(2), 194–218. https://doi.org/10.1080/1554477X.2020.1731280

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

Creamer, E. G. (2012). Effects of numeric representation of women on interest in engineering as a career. *American Journal of Engineering Education (AJEE)*, *3*(1), 1–12. https://doi.org/10.19030/ajee.v3i1.6884

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). Academic Press. https://doi.org/10.1016/S0065-2601(07)00002-0

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

DeBruine, L., & Jones, B. (2017). *Face Research Lab London Set (Version 3)* (p. 281699312 Bytes) [dataset]. [object Object]. https://doi.org/10.6084/M9.FIGSHARE.5047666.V3

Degner, J., Mangels, J., & Zander, L. (2019a). 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

Degner, J., Mangels, J., & Zander, L. (2019b). Visualizing gendered representations of male and female teachers using a reverse correlation paradigm. *Social Psychology*, *50*(4), 233–251. https://doi.org/10.1027/1864-9335/a000382

Devine, P. G. (1989). *Stereotypes and Prejudice: Their Automatic and Controlled Components*. 14.

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/10.1080/01650250344000226

Diekman, 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

Diekman, A. B., & Schneider, M. C. (2010). A social role theory perspective on gender gaps in political attitudes. *Psychology of Women Quarterly*, *34*(4), 486–497. https://doi.org/10.1111/j.1471-6402.2010.01598.x

Ditonto, T., & Mattes, K. (2018). Differences in Appearance-Based Trait Inferences for Male and Female Political Candidates. *Journal of Women, Politics & Policy*, *39*(4), 430–450. https://doi.org/10.1080/1554477X.2018.1506206

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2019). At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives. *Perspectives on Psychological Science*, *14*(2), 273–291. https://doi.org/10.1177/1745691617746796

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., & VAN Knippenberg, 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

Dowe, P. K. F. (2020). Resisting Marginalization: Black Women's Political Ambition and Agency. *PS: Political Science & Politics*, *53*(4), 697–702. https://doi.org/10.1017/S1049096520000554

Dunkle, J. H., & Francis, P. L. (1990). The role of facial masculinity/femininity in the attribution of homosexuality. *Sex Roles*, *23*(3–4), 157–167. https://doi.org/10.1007/BF00289863

Eagly, A. H. (1997). Sex differences in social behavior: Comparing social role theory and evolutionary psychology. *American Psychologist*, *52*(12), 1380–1383. https://doi.org/10.1037/0003-066X.52.12.1380.b

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Eagly, A. H., & Chin, J. L. (2010). Diversity and leadership in a changing world. *American Psychologist*, *65*(3), 216–224. https://doi.org/10.1037/a0018957

Eagly, A. H., & 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. (2002a). 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., & Karau, S. J. (2002b). 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. (2020a). 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., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020b). 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., & Sczesny, S. (2009). Stereotypes about women, men, and leaders: Have times changed? In M. Barreto, M. K. Ryan, & M. T. Schmitt (Eds.), *The glass ceiling in the 21st century: Understanding barriers to gender equality.* (pp. 21–47). American Psychological Association. https://doi.org/10.1037/11863-002

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 P. Van Lange, A. Kruglanski, & E. Higgins, *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 satisfier 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

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Farina, A., Fischer, E. H., Sherman, S., Smith, W. T., Groh, T., & Mermin, P. (1977). Physical attractiveness and mental illness. *Journal of Abnormal Psychology*, *86*(5), 510–517. https://doi.org/10.1037/0021-843X.86.5.510

Fink, B., & Penton-Voak, I. (2002). Evolutionary Psychology of Facial Attractiveness. *Current Directions in Psychological Science*, *11*(5), 154–158.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Bavel, J. J. V., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (2018). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins. In *Social Cognition*. Routledge.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002a). 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

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002b). 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

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002c). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Foos, P. W., & Clark, M. C. (2011). Adult Age and Gender Differences in Perceptions of Facial Attractiveness: Beauty is in the Eye of the Older Beholder. *The Journal of Genetic Psychology*, *172*(2), 162–175. https://doi.org/10.1080/00221325.2010.526154

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Gadassi, R., & Gati, I. (2009). The Effect of Gender Stereotypes on Explicit and Implicit Career Preferences. *The Counseling Psychologist*, *37*(6), 902–922. https://doi.org/10.1177/0011000009334093

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

Garcia-Retamero, R., & López-Zafra, E. (2006). Prejudice against Women in Male-congenial Environments: Perceptions of Gender Role Congruity in Leadership. *Sex Roles*, *55*(1–2), 51–61. https://doi.org/10.1007/s11199-006-9068-1

Geis, F. L. (1993). Self-fulfilling prophecies: A social psychological view of gender. In *The psychology of gender* (pp. 9–54). Guilford Press.

Gershon, S. A., & Lavariega Monforti, J. (2021). Intersecting campaigns: Candidate race, ethnicity, gender and voter evaluations. *Politics, Groups, and Identities*, *9*(3), 439–463. https://doi.org/10.1080/21565503.2019.1584752

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), 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., Adetoun, B., Osagie, J. E., Akande, A., Alao, A., Annetje, B., Willemsen, T. M., Chipeta, K., Dardenne, B.,

Dijksterhuis, A., Wigboldus, D., Eckes, T., Six-Materna, I., Expósito, F., … 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. (2021a). Narrow prototypes and neglected victims: Understanding perceptions of sexual harassment. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000260

Goh, J. X., Bandt-Law, B., Cheek, N. N., Sinclair, S., & Kaiser, C. R. (2021b). Narrow prototypes and neglected victims: Understanding perceptions of sexual harassment. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000260

Goodale, B. M., Alt, N. P., Lick, D. J., & Johnson, K. L. (2018). Groups at a glance: Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *Journal of Experimental Psychology: General*, *147*(11), 1660–1676. https://doi.org/10.1037/xge0000450

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3–25. https://doi.org/10.1037/0033-295X.109.1.3

Gundersen, A. B., & Kunst, J. R. (2019). Feminist ≠ Feminine? Feminist Women Are Visually Masculinized Whereas Feminist Men Are Feminized. *Sex Roles*, *80*(5), 291–309. https://doi.org/10.1007/s11199-018-0931-7

Haines, E. L., & Stroessner, S. J. (2019a). The role prioritization model: How communal men and agentic women can (sometimes) have it all. *Social and Personality Psychology Compass*, *13*(12), e12504. https://doi.org/10.1111/spc3.12504

Haines, E. L., & Stroessner, S. J. (2019b). The role prioritization model: How communal men and agentic women can (sometimes) have it all. *Social and Personality Psychology Compass*, *13*(12), e12504. https://doi.org/10.1111/spc3.12504

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2023). Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behavior Research Methods*, *55*(8), 3953–3964. https://doi.org/10.3758/s13428-022-01999-x

Hayes, D. (2011). When Gender and Party Collide: Stereotyping in Candidate Trait Attribution. *Politics & Gender*, *7*(02), 133–165. https://doi.org/10.1017/S1743923X11000055

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014a). 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., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014b). 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

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology, 92*(1), 81–92. https://doi.org/10.1037/0021-9010.92.1.81

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

Hicks, H. M. (2022). Intersectional Stereotyping in Media Coverage: The Case of Stacey Abrams Versus Stacey Evans in Georgia. *Journal of Women, Politics & Policy, 43*(1), 95–106. https://doi.org/10.1080/1554477X.2022.2005388

Ho, A. K., Sidanius, J., Kteily, N. S., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO$_7$ scale. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000033

Hoffman, C., & Hurst, N. (1990). Gender stereotypes: Perception or rationalization? *Journal of Personality and Social Psychology, 58*(2), 197–208. https://doi.org/10.1037/0022-3514.58.2.197

Hollinger, D. A. (2005). The One Drop Rule & the One Hate Rule. *Daedalus*, *134*(1), 18–28.

Hong, Y., Chua, K.-W., & Freeman, J. B. (2024). Reducing Facial Stereotype Bias in Consequential Social Judgments: Intervention Success With White Male Faces. *Psychological Science*, *35*(1), 21–33. https://doi.org/10.1177/09567976231215238

Hoss, R. A., Ramsey, J. L., Griffin, A. M., & Langlois, J. H. (2005). The Role of Facial Attractiveness and Facial Masculinity/Femininity in Sex Classification of Faces. *Perception*, *34*(12), 1459–1474.

Hsu, N., Badura, K. L., Newman, D. A., & Speach, M. E. P. (2021). Gender, "masculinity," and "femininity": A meta-analytic review of gender differences in agency and communion. *Psychological Bulletin*, *147*(10), 987–1011. https://doi.org/10.1037/bul0000343.supp

Hu, Y., Abbasi, N. ul H., Zhang, Y., & Chen, H. (2018). The Effect of Target Sex, Sexual Dimorphism, and Facial Attractiveness on Perceptions of Target Attractiveness and Trustworthiness. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00942

Hughes, S. M., & Bremme, R. (2011). The effects of facial symmetry and sexually-dimorphic facial proportions on assessments of sexual orientation. *Journal of Social, Evolutionary, and Cultural Psychology*, *5*(4), 214–230. https://doi.org/10.1037/h0099261

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

Iyengar, S., & Krupenkin, M. (2018). The Strengthening of Partisan Affect. *Political Psychology*, *39*(S1), 201–218. https://doi.org/10.1111/pops.12487

Jackson, L. A., Sullivan, L. A., & Hymes, J. S. (1987). Gender, Gender Role, and Physical Appearance. *The Journal of Psychology*, *121*(1), 51–56. https://doi.org/10.1080/00223980.1987.9712642

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

James, E. A., Jenkins, S., & Watkins, C. D. (2018). Negative Effects of Makeup Use on Perceptions of Leadership Ability Across Two Ethnicities. *Perception*, *47*(5), 540–549. https://doi.org/10.1177/0301006618763263

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012a). 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

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012b). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*(1), 116–131. https://doi.org/10.1037/a0025335

Joshi, M. P., Lloyd, E. P., Diekman, A. B., & Hugenberg, K. (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

Joshi, M. P., Lloyd, E. P., Diekman, A. B., & Hugenberg, K. (2023). In the face of opportunities: Facial structures of scientists shape expectations of STEM environments. *Personality and Social Psychology Bulletin*, *49*(5), 673–691. https://doi.org/10.1177/01461672221077801

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Jost, J. T., Nosek, B. A., & Gosling, S. D. (2008). Ideology: Its Resurgence in Social, Personality, and Political Psychology. *Perspectives on Psychological Science*, *3*(2), 126–136. https://doi.org/10.1111/j.1745-6916.2008.00070.x

Kahn, S., & Ginther, D. K. (2017). *Women and Stem* (SSRN Scholarly Paper ID 2988746). Social Science Research Network. https://papers.ssrn.com/abstract=2988746

Karl, K., Peluchette, J. V., & Dawson, G. (2022). Why So Much Ado About a Hairdo? Examining How the Hair Choices of Black Women Vary by Occupation. *Journal of Business Diversity*, *22*(1), Article 1. https://articlearchives.co/index.php/JBD/article/view/1907

Keith, M. G., & McKay, A. S. (2024). Too Anecdotal to Be True? Mechanical Turk Is Not All Bots and Bad Data: Response to Webb and Tangney (2022). *Perspectives on Psychological Science*, 17456916241234328. https://doi.org/10.1177/17456916241234328

Kleider-Offutt, H., Meacham, A. M., Branum-Martin, L., & Capodanno, M. (2021). What's in a face? The role of facial features in ratings of dominance, threat, and stereotypicality. *Cognitive Research: Principles and Implications*, *6*(1), 53. https://doi.org/10.1186/s41235-021-00319-9

Knuckey, J., & Kim, M. (2015). Racial Resentment, Old-Fashioned Racism, and the Vote Choice of Southern and Nonsouthern Whites in the 2012 U.S. Presidential Election*. *Social Science Quarterly*, *96*(4), 905–922. https://doi.org/10.1111/ssqu.12184

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*(3), 165–182. https://doi.org/10.1007/BF02281954

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

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 1–11. https://doi.org/10.1037/a0034309

Lick, D. J., & Johnson, K. L. (2013). Fluency of visual processing explains prejudiced evaluations following categorization of concealable identities. *Journal of Experimental Social Psychology, 49*(3), 419–425. https://doi.org/10.1016/j.jesp.2013.01.001

Lick, D. J., & Johnson, K. L. (2014). Recalibrating gender perception: Face aftereffects and the perceptual underpinnings of gender-related biases. *Journal of Experimental Psychology: General, 143*(3), 1259–1276. https://doi.org/10.1037/a0034516

Lick, D. J., & Johnson, K. L. (2015). The Interpersonal Consequences of Processing Ease: Fluency as a Metacognitive Foundation for Prejudice. *Current Directions in Psychological Science*, *24*(2), 143–148. https://doi.org/10.1177/0963721414558116

Lick, D. J., & Johnson, K. L. (2018). Facial Cues to Race and Gender Interactively Guide Age Judgments. *Social Cognition*, *36*(5), 497–516. https://doi.org/10.1521/soco.2018.36.5.497

Liebenow, H. A., Boucher, K. L., & Cassidy, B. S. (2022). Understanding Evaluations of Kamala Harris in 2020: Political Ideology Qualifies Perceived Communality Effects When Communal Cues Are Present. *Psychology of Women Quarterly*, *46*(4), 399–419. https://doi.org/10.1177/03616843221104383

Liebenow, H. A., Boucher, K. L., & Cassidy, B. S. (2024). Trait inferences from the "big two" produce gendered expectations of facial features. *Journal of Experimental Social Psychology*, *111*, 104585. https://doi.org/10.1016/j.jesp.2023.104585

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

Little, A., Jones, B., & DeBruine, L. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 1638–1659. https://doi.org/10.1098/rstb.2010.0404

Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an Agentic Black Woman Get Ahead? The Impact of Race and Interpersonal Dominance on Perceptions of Female Leaders. *Psychological Science*, *23*(4), 354–358. https://doi.org/10.1177/0956797611428079

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. WY., Kidd, C., Marcelle, E., … Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron*, *109*(13), 2047–2074. https://doi.org/10.1016/j.neuron.2021.06.002

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

Ma, D. S., Hohl, D., & Kantner, J. (2021). The politics of identity: The unexpected role of political orientation on racial categorizations of Kamala Harris. *Analyses of Social Issues and Public Policy*, asap.12257. https://doi.org/10.1111/asap.12257

Mallinas, S. R., Crawford, J. T., & Cole, S. (2018). Political Opposites Do Not Attract: The Effects of Ideological Dissimilarity on Impression Formation. *Journal of Social and Political Psychology*, *6*(1), Article 1. https://doi.org/10.5964/jspp.v6i1.747

Marcinkowska, U. M., Kozlov, M. V., Cai, H., Contreras-Garduño, J., Dixson, B. J., Oana, G. A., Kaminski, G., Li, N. P., Lyons, M. T., Onyishi, I. E., Prasai, K., Pazhoohi, F., Prokop, P., Rosales Cardozo, S. L., Sydney, N., Yong, J. C., & Rantala, M. J. (2014). Cross-cultural variation in men's preference for sexual dimorphism in women's faces. *Biology Letters*, *10*(4), 20130850. https://doi.org/10.1098/rsbl.2013.0850

Martin, A. E., & Slepian, M. L. (2020a). 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

Martin, A. E., & Slepian, M. L. (2020b). 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

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*(5), 806–816. https://doi.org/10.1002/ejsp.445

McGorray, E. L., Emery, L. F., Garr-Schultz, A., & Finkel, E. J. (2023). "Mostly White, heterosexual couples": Examining demographic diversity and reporting practices in relationship science research samples. *Journal of Personality and Social Psychology*, *125*(2), 316–344. https://doi.org/10.1037/pspi0000417

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The Morphometrics of "Masculinity" in Human Faces. *PLOS ONE*, *10*(2), e0118374. https://doi.org/10.1371/journal.pone.0118374

Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology*, *75*(4), 901–916. https://doi.org/10.1037/0022-3514.75.4.901

Morgan, G., Mullen, E., & Skitka, L. (2010). When Values and Attributions Collide: Liberals' and Conservatives' Values Motivate Attributions for Alleged Misdeeds. *Personality & Social Psychology Bulletin*, *36*, 1241–1254. https://doi.org/10.1177/0146167210380605

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

Motro, D., Evans, J. B., Ellis, A. P. J., & Benson, L. (2021). Race and reactions to women's expressions of anger at work: Examining the effects of the "angry Black woman" stereotype. *Journal of Applied Psychology*. https://doi.org/10.1037/apl0000884

Myers, D., & Levy, M. (2018). Racial Population Projections and Reactions to Alternative News Accounts of Growing Diversity. *The ANNALS of the American Academy of Political and Social Science*, *677*(1), 215–228. https://doi.org/10.1177/0002716218766294

Nelson, A., & Piatak, J. (2021). Intersectionality, Leadership, and Inclusion: How Do Racially Underrepresented Women Fare in the Federal Government? *Review of Public Personnel Administration*, *41*(2), 294–318. https://doi.org/10.1177/0734371X19881681

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing Hidden Gender Biases in Competence Impressions of Faces. *Psychological Science*, *30*(1), 65–79. https://doi.org/10.1177/0956797618813092

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (n.d.). *Gender Biases in Impressions From Faces: Empirical Studies and Computational Models*.

Okimoto, T. G., & Brescoll, V. L. (2010). The Price of Power: Power Seeking and Backlash Against Female Politicians. *Personality and Social Psychology Bulletin*, *36*(7), 923–936. https://doi.org/10.1177/0146167210371949

Oliveira, M., Garcia-Marques, T., 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

Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2020). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two? *European Journal of Social Psychology*, *50*(2), 463–483. https://doi.org/10.1002/ejsp.2618

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

Olszanowski, M., Kaminska, O. K., & Winkielman, P. (2018). Mixed matters: Fluency impacts trust ratings when faces range on valence but not on motivational implications. *Cognition and Emotion*, *32*(5), 1032–1051. https://doi.org/10.1080/02699931.2017.1386622

Oosterhof, N. N., & Todorov, A. (2008). The Functional Basis of Face Evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Oosterhof, N., & Todorov, A. (2008). The Functional Basis of Face Evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 11087–11092. https://doi.org/10.1073/pnas.0805664105

Pauker, K., Carpinella, C., Meyers, C., Young, D. M., & Sanchez, D. T. (2018). The Role of Diversity Exposure in Whites' Reduction in Race Essentialism Over Time. *Social Psychological and Personality Science*, *9*(8), 944–952. https://doi.org/10.1177/1948550617731496

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology*, *46*(2), 367–374. https://doi.org/10.1016/j.jesp.2009.11.001

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L., & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*(6696), Article 6696. https://doi.org/10.1038/29772

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

Piatek-Jimenez, K., Cribbs, J., & Gill, N. (2018). College students' perceptions of gender stereotypes: Making connections to the underrepresentation of women in STEM fields. *International Journal of Science Education*, *40*(12), 1432–1454. https://doi.org/10.1080/09500693.2018.1482027

Piston, S. (2010). How Explicit Racial Prejudice Hurt Obama in the 2008 Election. *Political Behavior*, *32*(4), 431–451. https://doi.org/10.1007/s11109-010-9108-y

Pleasants, 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*. https://dash.harvard.edu/handle/1/3207711

Prentice, D. A., & Carranza, E. (2002a). What women and men should Be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, *26*(4), 269–281. https://doi.org/10.1111/1471-6402.t01-1-00066

Prentice, D. A., & Carranza, E. (2002b). What Women and Men Should Be, Shouldn't be, are Allowed to be, and don't Have to Be: The Contents of Prescriptive Gender Stereotypes. *Psychology of Women Quarterly*, *26*(4), 269–281. https://doi.org/10.1111/1471-6402.t01-1-00066

Quittkat, H. L., Hartmann, A. S., Düsing, R., Buhlmann, U., & Vocks, S. (2019). Body Dissatisfaction, Importance of Appearance, and Body Appreciation in Men and Women Over the Lifespan. *Frontiers in Psychiatry*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00864

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

Ratliff, K. A., Redford, L., Conway, J., & Smith, C. T. (2019). Engendering support: Hostile sexism predicts voting for Donald Trump over Hillary Clinton in the 2016 U.S. presidential election. *Group Processes & Intergroup Relations*, *22*(4), 578–593. https://doi.org/10.1177/1368430217741203

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

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(suppl_1), S93–S95. https://doi.org/10.1098/rsbl.2003.0023

Ritter, B. A., & Yoder, J. D. (2004). Gender Differences in Leader Emergence Persist Even for Dominant Women: An Updated Confirmation of Role Congruity Theory. *Psychology of Women Quarterly*, *28*(3), 187–193. https://doi.org/10.1111/j.1471-6402.2004.00135.x

Roberts, S. O. (2024). Dealing With Diversity in Psychology: Science and Ideology. *Perspectives on Psychological Science*, *19*(3), 590–601. https://doi.org/10.1177/17456916241240743

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial Inequality in Psychological Research: Trends of the Past and Recommendations for the Future. *Perspectives on Psychological Science*, *15*(6), 1295–1309. https://doi.org/10.1177/1745691620927709

Rogowski, J. C., & Sutherland, J. L. (2016). How Ideology Fuels Affective Polarization. *Political Behavior*, *38*(2), 485–508. https://doi.org/10.1007/s11109-015-9323-7

Rosette, A. S., Koval, C., Ma, A., & Livingston, R. B. (2016). *Race matters for women leaders: Intersectional effects on agentic deficiencies and penalties*. https://doi.org/10.1016/J.LEAQUA.2016.01.008

Rosette, A. S., & Tost, L. P. (2010a). 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

Rosette, A. S., & Tost, L. P. (2010b). 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

Rothwell, V., Hodson, G., & Prusaczyk, E. (2019). Why Pillory Hillary? Testing the endemic sexism hypothesis regarding the 2016 U.S. election. *Personality and Individual Differences*, *138*, 106–108. https://doi.org/10.1016/j.paid.2018.09.034

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. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629–645. https://doi.org/10.1037/0022-3514.74.3.629

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers.

Journal of Personality and Social Psychology, *77*(5), 1004–1010. https://doi.org/10.1037/0022-3514.77.5.1004

Rudman, L. A., & Glick, P. (2001a). 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., & Glick, P. (2001b). 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., & Glick, P. (2001c). 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

Rudman, L. A., & Phelan, J. E. (2008a). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, *28*, 61–79. https://doi.org/10.1016/j.riob.2008.04.003

Rudman, L. A., & Phelan, J. E. (2008b). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, *28*, 61–79. https://doi.org/10.1016/j.riob.2008.04.003

Rudman, L., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of*

Personality and Social Psychology, *77*, 1004–1010. https://doi.org/10.1037/0022-3514.77.5.1004

Rudman, L., & Goodwin, S. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*, 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Rule, N. O., & Ambady, N. (2008). The Face of Success: Inferences From Chief Executive Officers' Appearance Predict Company Profits. *Psychological Science*, *19*(2), 109–111. https://doi.org/10.1111/j.1467-9280.2008.02054.x

Rule, N. O., 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

Sanchez-Hucles, J. V., & Davis, D. D. (2010). Women and women of color in leadership: Complexity, identity, and intersectionality. *The American Psychologist*, *65*(3), 171–181. https://doi.org/10.1037/a0017459

Scheuer, C.-L., & Loughlin, C. (2020). Could the aging workforce reduce the agency penalty for female leaders? Re-examining the think manager–think male stereotype. *Journal of Management & Organization*, *26*(1), 29–51. https://doi.org/10.1017/jmo.2018.41

Schneider, B., Klager, C., Chen, I.-C., & Burns, J. (2016). Transitioning Into Adulthood: Striking a Balance Between Support and Independence. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 106–113. https://doi.org/10.1177/2372732215624932

Schneider, M. C., & Bos, A. L. (2019). The Application of Social Role Theory to the Study of Gender in Politics. *Political Psychology*, *40*(S1), 173–213. https://doi.org/10.1111/pops.12573

Schneider, M. C., Bos, A. L., & DiFilippo, M. (2021a). Gender Role Violations and Voter Prejudice: The Agentic Penalty Faced by Women Politicians. *Journal of Women, Politics & Policy*, *0*(0), 1–17. https://doi.org/10.1080/1554477X.2021.1981095

Schneider, M. C., Bos, A. L., & DiFilippo, M. (2021b). Gender role violations and voter prejudice: The agentic penalty faced by women politicians. *Journal of Women, Politics & Policy*, 1–17. https://doi.org/10.1080/1554477X.2021.1981095

Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Processes*, *101*(2), 230–242. https://doi.org/10.1016/j.obhdp.2006.06.002

Seamon, J. G., Williams, P. C., Crowley, M. J., Kim, I. J., Langer, S. A., Orne, P. J., & Wishengrad, D. L. (1995). The mere exposure effect is based on implicit memory: Effects of stimulus type, encoding conditions, and number of exposures on recognition and affect judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 711–721. https://doi.org/10.1037/0278-7393.21.3.711

Shen, X., & Ferguson, M. J. (2021a). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, *97*, 104219. https://doi.org/10.1016/j.jesp.2021.104219

Shen, X., & Ferguson, M. J. (2021b). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, *97*, 104219. https://doi.org/10.1016/j.jesp.2021.104219

Sherman, G. D., & Haidt, J. (2011). Cuteness and Disgust: The Humanizing and Dehumanizing Effects of Emotion. *Emotion Review*, *3*(3), 245–251. https://doi.org/10.1177/1754073911402396

Shiramizu, V. K. M., Batres, C., Jones, A. L., Stern, J., DeBruine, L. M., & Jones, B. (2024). *Who looks competent for what? Context-sensitivity in the predictors of competence judgments of face*. https://doi.org/10.31234/osf.io/eh52z

Silván-Ferrero, M. del P., & Bustillos López, A. (2007). Benevolent Sexism Toward Men and Women: Justification of the Traditional System and Conventional Gender Roles in Spain. *Sex Roles*, *57*(7–8), 607–614. https://doi.org/10.1007/s11199-007-9271-8

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

Smith, J. L., Paul, D., & Paul, R. (2007). No Place for a Woman: Evidence for Gender Bias in Evaluations of Presidential Candidates. *Basic and Applied Social Psychology*, *29*(3), 225–233. https://doi.org/10.1080/01973530701503069

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What Is Typical Is Good: The Influence of Face Typicality on Perceived Trustworthiness. *Psychological Science*, *26*(1), 39–47. https://doi.org/10.1177/0956797614554955

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

Steiner, T. G., Vescio, T. K., & Adams, R. B. (2022). The effect of gender identity and gender threat on self-image. *Journal of Experimental Social Psychology*, *101*, 104335. https://doi.org/10.1016/j.jesp.2022.104335

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015a). 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

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015b). 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

Tagai, K., Ohtaka, H., & Nittono, H. (2016). Faces with Light Makeup Are Better Recognized than Faces with Heavy Makeup. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00226

Tagai, K., Shimakura, H., Isobe, H., & Nittono, H. (2017). The light-makeup advantage in facial processing: Evidence from event-related potentials. *PLOS ONE*, *12*(2), e0172489. https://doi.org/10.1371/journal.pone.0172489

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, *223*(5), 96–103.

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

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

Tresh, F., Steeden, B., Randsley de Moura, G., Leite, A. C., Swift, H. J., & Player, A. (2019). Endorsing and Reinforcing Gender and Age Stereotypes: The Negative Effect on Self-Rated Leadership Potential for Women and Older Workers. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.00688

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

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

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

Villavisanis, D. F., Blum, J. D., Liou, L., Cho, D. Y., Swanson, J. W., Bartlett, S. P., & Taylor, J. A. (2023). Real Face Analysis of Masculinity-Femininity, Facial Ratios, and Attractiveness. *Annals of Plastic Surgery*, *91*(2), 287. https://doi.org/10.1097/SAP.0000000000003616

Walker, M., & Wänke, M. (2017a). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PLOS ONE*, *12*(10), e0181306. https://doi.org/10.1371/journal.pone.0181306

Walker, M., & Wänke, M. (2017b). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PLOS ONE*, *12*(10), e0181306. https://doi.org/10.1371/journal.pone.0181306

Webb, M. A., & Tangney, J. P. (2022). Too Good to Be True: Bots and Bad Data From Mechanical Turk. *Perspectives on Psychological Science*, 17456916221120027. https://doi.org/10.1177/17456916221120027

Wen, F., Zuo, B., Ma, S., Xu, Y., Coley, J. D., & Wang, Y. (2020a). Do we see masculine faces as competent and feminine faces as warm? Effects of sexual dimorphism on facial perception. *Evolutionary Psychology*, *18*(4), 147470492098064. https://doi.org/10.1177/1474704920980642

Wen, F., Zuo, B., Ma, S., Xu, Y., Coley, J. D., & Wang, Y. (2020b). Do we see masculine faces as competent and feminine faces as warm? Effects of sexual dimorphism on facial perception. *Evolutionary Psychology*, *18*(4), 147470492098064. https://doi.org/10.1177/1474704920980642

West, E. A., & Iyengar, S. (2020). Partisanship as a Social Identity: Implications for Polarization. *Political Behavior*. https://doi.org/10.1007/s11109-020-09637-y

Wetherell, G. A., Brandt, M. J., & Reyna, C. (2013). Discrimination Across the Ideological Divide: The Role of Value Violations and Abstract Values in Discrimination by Liberals and Conservatives. *Social Psychological and Personality Science*, *4*(6), 658–667. https://doi.org/10.1177/1948550613476096

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.

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, *142*(2), 165–197. https://doi.org/10.1037/bul0000039

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., Remedios, J. D., & Rule, N. O. (2017). Interactive Effects of Obvious and Ambiguous Social Categories on Perceptions of Leadership: When Double-Minority Status May Be Beneficial. *Personality and Social Psychology Bulletin*, *43*(6), 888–900. https://doi.org/10.1177/0146167217702373

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, *17*(9), 799–806. https://doi.org/10.1111/j.1467-9280.2006.01785.x

Winter, N. J. G. (2010). Masculine Republicans and Feminine Democrats: Gender and Americans' Explicit and Implicit Images of the Political Parties. *Political Behavior*, *32*(4), 587–618. https://doi.org/10.1007/s11109-010-9131-z

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

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. https://doi.org/10.1037/pspi0000160.supp

Yang, T., Chen, H., Hu, Y., Zheng, Y., & Wang, W. (2015). Preferences for sexual dimorphism on attractiveness levels: An eye-tracking study. *Personality and Individual Differences*, *77*, 179–185. https://doi.org/10.1016/j.paid.2014.12.005

Yee, A. H. (1973). Are the schools a "feminized society"? *Educational Leadership*, *31*(2), 128–133.

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

APPENDIX A: SUPPPLEMENTAL MATERIAL FOR CHAPTER II

Models Regressing Evaluations of Harris on Participant Characterization Variables

Table A1

Regression summaries for evaluations of Kamala Harris and her candidacy by participant characterization variables (all standardized around the overall mean) in Study 1 (N = 165).

	a. Impression Positivity			**b. Support**			**c. Expected Success**			**d. Voting Likelihood**		
	b (SE)	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*
Ideology	1.40 (0.13)	10.28	<.001	1.68 (.15)	11.34	<.001	1.31 (0.12)	10.95	<.001	1.50 (0.15)	10.10	<.001
Hostile sexism	-0.05 (0.17)	0.32	.75	-0.07 (0.19)	0.40	.69	-0.21 (0.15)	1.40	.16	-0.24 (0.19)	1.27	.21
Benevolent sexism	0.46 (.16)	2.96	.004	0.37 (0.17)	2.21	.03	0.22 (0.14)	.14	.11	0.37 (0.17)	2.20	.03
Non-transcendent gender beliefs	-0.01 (0.14)	-0.10	.92	-0.09 (0.15)	0.60	.55	-0.06 (0.12)	0.48	.63	-0.21 (0.15)	1.41	.16
Gender-linked beliefs	-0.04 (0.19)	0.20	.85	-0.01 (0.21)	0.05	.96	0.03 (0.17)	0.16	.87	0.07 (0.21)	0.35	.73

Table A2

Regression summaries for evaluations of Kamala Harris and her candidacy by participant characterization variables (all standardized around the overall mean) in Study 2 (N = 263).

	a. Impression Positivity			**b. Support**			**c. Expected Success**			**d. Voting Likelihood**		
	b (SE)	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*	*b (SE)*	*t*	*p*
Ideology	1.19 (0.09)	11.99	<.001	1.43 (.11)	12.73	<.001	1.01 (0.08)	11.06	<.001	1.44 (0.09)	13.57	<.001
Hostile sexism	-0.56 (0.12)	4.55	<.001	-0.72 (0.14)	5.19	<.001	-0.63 (0.11)	5.59	<.001	-0.75 (0.13)	5.72	<.001
Benevolent sexism	0.19 (.11)	1.79	.08	0.22 (0.12)	1.81	.07	0.25 (0.10)	2.58	.01	0.21 (0.12)	1.86	.06
Non-transcendent gender beliefs	-0.12 (0.10)	1.15	.25	-0.25 (0.15)	2.13	.03	-0.26 (0.09)	2/73	.007	-0.27 (0.11)	2.39	.02
Gender-linked beliefs	0.30 (0.13)	2.24	.03	0.31 (0.15)	1.97	.05	0.13 (0.12)	1.06	.29	0.29 (0.14)	2.02	.04

Having a more liberal political ideology consistently had the strongest positive relation with evaluations across studies. However, benevolent sexism had some positive relations with evaluations across studies and more positive relations emerged overall in Study 2 than in Study 1. More emergent positive relations with evaluations overall in Study 2 relative to Study 1 was likely due to the increased sample size in Study 2. That is, relations with, for example, hostile sexism, might have been more readily detectable in a larger sample. Although the emergent relations across studies do not suggest that ideology is the *only* positive predictor of evaluations, they do suggest that ideology provides strongly consistent relations beyond relations with other variables that are more face valid in underlying sexism and beliefs about gender roles. We thus interpreted our findings with regard to political ideology from a social identity-based perspective.

Models Aggregating Across Evaluations

We made an a priori decision to run a multivariate regression followed by univariate regressions for each of the four evaluations. However, we also ran a univariate regression after aggregating across the four evaluations. We quantified the average evaluation of Kamala Harris for each participant by creating an average of the four evaluations (impression positivity, support, expected success, and voting likelihood).

Study 1

We regressed the average evaluation of Harris on standardized composite communality inferences, composite agency inferences, composite ideology scores, and their interactions. The model was significant, $F(7, 157) = 110.90$, $p < .001$, $R^2 = .82$. See Table 3 for coefficient information. The significant interaction between perceived communality and political ideology showed that perceived communality had a stronger positive relationship with expected success for more conservative, $b = 1.61$, $SE = 0.11$, $t = 14.58$, $p < .001$, than liberal, $b = 1.27$, $SE = 0.15$, $t = 8.35$, $p < .001$, participants.

Table 3

Regression summary for average evaluation of Kamala Harris by gendered trait inferences and political ideology in Study 1 (N = 165).

	b (SE)	*t*	*p*
Communality	1.44 (0.10)	14.09	<.001
Agency	0.20 (0.11)	1.80	.07
Ideology	0.45 (0.10)	4.48	<.001
Communality * Agency	-0.07 (0.08)	0.92	.36
Communality * Ideology	-0.17 (0.09)	1.99	.048
Agency * Ideology	0.13 (0.13)	1.00	.32
Communality * Agency * Ideology	-0.05 (0.08)	0.71	.48

Study 2

We regressed the average evaluation of Harris on standardized composite communality inferences, composite agency inferences, composite ideology scores, role information (agentic = -1, communal = 1) and their interactions. The model was significant, $F(15, 247) = 59.06$, $p < .001$, $R^2 = .77$. See Table 4 for coefficient information.

The Communion × Ideology × Description interaction showed that when the conveyed information was more agentic, communality effects similarly emerged across more conservative, $b = 1.29$, $SE = 0.12$, $t = 10.67$, $p < .001$, and liberal, $b = 1.42$, $SE = 0.21$, $t = 6.77$, $p < .001$, participants, $b = -0.12$, $SE = 0.21$, $t = 0.56$, $p = 0.57$. When conveyed information was more communal, communality effects were stronger for more conservative, $b = 1.36$, $SE = 0.16$, $t = 8.54$, $p < .001$, than liberal, $b = 0.66$, $SE = 0.21$, $t = 3.12$, $p = .002$, participants, $b = 0.70$, $SE = 0.26$, $t = 2.74$, $p = 0.01$.

The Agency × Communion × Ideology interaction showed that when Harris was perceived as less agentic, similar communality effects emerged across more liberal, $b = 1.31$, $SE = 0.20$, $t = 6.57$, $p < .001$, and more conservative, $b = 0.96$, $SE = 0.13$, $t = 7.57$, participants, $b = -0.35$, $SE = 0.22$, $t = 1.58$, $p = .12$. When more agentic, communality effects were stronger for more conservative, $b = 1.70$, $SE = 0.15$, $t = 11.13$, $p < .001$, than liberal, $b = 0.77$, $SE = 0.17$, $t = 4.43$, $p < .001$, participants, $b = 0.93$, $SE = 0.22$, $t = 4.23$, $p < .001$.

Table 4

Regression summary for average evaluation of Kamala Harris by gendered trait inferences, political ideology, and description in Study 2 (N = 263).

	b (SE)	*t*	*p*
Communality	1.18 (0.10)	12.32	< .001
Agency	0.42 (0.10)	4.19	< .001
Ideology	0.87 (0.09)	10.01	< .001
Description	-0.05 (0.08)	0.59	.56
Communality * Agency	0.05 (0.08)	0.69	.49
Communality * Ideology	-0.15 (0.08)	1.76	.08
Agency * Ideology	-0.16 (0.10)	1.63	.10
Communality * Description	-0.17 (0.10)	1.79	.08
Agency * Description	0.05 (0.10)	0.53	.59
Ideology * Description	0.08 (0.09)	0.93	.35
Communality * Agency * Ideology	-0.32 (0.07)	4.38	< .001
Communality * Agency * Description	0.03 (0.08)	0.39	.70
Communality * Ideology * Description	-0.21 (0.08)	2.48	.01
Agency * Ideology * Description	0.12 (0.10)	1.23	.22
Communality * Agency * Ideology * Description	-0.02 (0.07)	0.24	.81

Factor Analyses for Studies 1 and 2

For Study 1, a confirmatory factor analysis using a varimax rotation on the four evaluations showed that one factor was sufficient to explain the dataset, $\chi^2 = 2.81$, $p = .25$. For Study 2, a confirmatory factor analysis using a varimax rotation on the four evaluations showed that one factor was not sufficient to explain the dataset, $\chi^2 = 16.70$, $p < .001$.

Additional Statistics for Discussion Sections for Studies 1 and 2

Study 1

As expected, the positive relation between perceived communality and expected success was attenuated among more liberal than conservative perceivers. This pattern was due to more liberal than conservative perceivers expecting more success when Harris was perceived as less, $t(157) = 8.36$, $p < .001$, but not more, $t(157) = 0.50$, $p = .62$, communal.

Unexpectedly, perceived agency and ideology interacted to affect expected success. This pattern differed from the interaction between perceived communality and ideology. Here, the positive relation between perceived agency and expected success was stronger for more liberal than conservative perceivers. This pattern was due to more liberal relative to conservative perceivers expecting more success when Harris was perceived as more, $t(157) = 5.39$, $p < .001$, but not less, $t(157) = 0.20$, $p = .84$, agentic.

Study 2

Indeed, across these evaluations, more liberal than conservative perceivers evaluated Harris positively to a greater extent when she was perceived as less, ts > 5.26, ps < .001, than more, ts > 1.86, ps < .06, communal.

Like Study 1, more liberal relative to conservative perceivers expected success to a greater extent when perceiving Harris as less, $t(247) = 6.63$, $p < .001$, than more, $t(247) = 2.94$, $p = .004$, communal.

APPENDIX B: SUPPPLEMENTAL MATERIAL FOR CHAPTER III

Chicago Face Database Norms

Participants responded to all ratings (except age) on a 1–7 Likert scale (1 = Not at all, 7 = Extremely).

Experiment 1 (WM-204 & WF-208; WM-029 & WF-242)

WM-204: Masculinity = 5.37; Femininity = 1.48; Age = 35.13; Attractiveness = 3.2

WF-208: Masculinity = 1.44; Femininity = 5.07; Age = 26.44; Attractiveness = 3.96

WM-029: Masculinity = 5.21; Femininity = 1.61; Age = 28.59; Attractiveness = 4.59

WF-242: Masculinity = 1.32; Femininity = 5.54; Age = 27.29; Attractiveness = 4.75

Experiments 2-3

Target	Race	Gender	Age	Attractive	Masculine	Feminine
WF-006	W	F	24.33684	3.478723	2.031915	4.606383
WF-011	W	F	24.04167	4.129032	1.702128	5.138298
WF-014	W	F	36.1	3.280899	2.370787	4.204545
WF-016	W	F	25.41111	3.707865	2.488636	4.52809
WF-017	W	F	23.30337	3.280899	2.404494	4.359551
WF-020	W	F	26.02222	3.325843	2.352273	4.382022
WF-031	W	F	23.23529	4.121951	1.843373	4.952381
WF-033	W	F	32.0404	3.393939	2.242424	4.515152
WF-035	W	F	24.76136	3.390805	2.574713	4.126437
WF-038	W	F	28.72283	3.315217	2.445652	4.293478
WF-039	W	F	23.7234	3.364583	2.03125	4.59375
WF-201	W	F	27	3.818182	2.545455	4.545455
WF-202	W	F	27.96552	3.62069	1.965517	4.896552
WF-213	W	F	25.65517	3.241379	2.103448	4.142857
WF-215	W	F	30.64	3.28	2.24	4.25
WF-217	W	F	32.88462	3.307692	1.692308	4.6
WF-218	W	F	28.96552	3.62069	1.827586	4.655172
WF-223	W	F	23.26923	3.384615	1.615385	5
WF-227	W	F	26.92593	3.769231	1.925926	4.666667
WF-228	W	F	36.23077	3.692308	1.807692	5.192308
WF-230	W	F	22.76	3.64	1.64	5.16
WF-234	W	F	22.62963	4.111111	1.185185	5.333333
WF-244	W	F	26.63636	3.818182	1.5	5.363636
WF-245	W	F	35.71429	3.428571	1.666667	4.607143
WF-246	W	F	42.82143	3.535714	1.464286	5.071429
WF-247	W	F	26.41667	3.25	1.583333	4.826087
WF-252	W	F	26.46429	3.785714	1.392857	5.107143
WM-003	W	M	23.34524	3.682927	4.891566	1.915663
WM-006	W	M	25.64535	3.511628	4.476744	2.581395
WM-009	W	M	23.69565	4.076087	4.728261	2.032609
WM-011	W	M	30.98864	3.170455	4.25	2.114943
WM-014	W	M	33.71429	3.47619	4.869048	1.940476
WM-015	W	M	23.37209	3.511628	4.302326	2.313953
WM-016	W	M	30.39535	3.255814	5.209302	1.406977
WM-024	W	M	20.81707	3.646341	3.95122	2.219512
WM-028	W	M	24.92683	3.2	4.2625	2.2
WM-029	W	M	28.58621	4.586207	5.206897	1.609195
WM-033	W	M	26.59341	3.846154	4.555556	1.78022
WM-200	W	M	19.5	3.269231	3.846154	1.730769
WM-203	W	M	29.25	3.5	4.869565	1.333333
WM-204	W	M	35.13333	3.2	5.366667	1.482759
WM-205	W	M	29.36	3.8	5	1.541667
WM-207	W	M	26.58333	4.125	5.208333	1.333333
WM-209	W	M	20.39286	3.535714	3.607143	2.142857
WM-211	W	M	27.7931	3.464286	4.607143	1.758621
WM-212	W	M	30.11111	3.703704	4.962963	1.814815
WM-213	W	M	26.36	3.708333	4.16	1.92
WM-214	W	M	25.12	3.12	3.666667	2.25
WM-230	W	M	21.60714	3.142857	3.821429	1.851852
WM-231	W	M	20.58621	3.642857	4.241379	1.862069
WM-235	W	M	20.75	3.357143	4.321429	1.571429
WM-238	W	M	24.81818	3.272727	4.590909	1.5
WM-240	W	M	34.38462	3.346154	4.5	1.92
WM-242	W	M	23.06897	4.034483	4.214286	1.758621
WM-243	W	M	31.44	3.16	4.652174	1.76
WM-245	W	M	26.27586	3.357143	4.62069	1.344828
WM-257	W	M	30.03704	3.740741	4.346154	1.814815

FaceGen

Dataset: https://facegen.com/download/scanDemographicsForPublic.zip

Offset Linear Controls: https://facegen.com/dl/sdk/doc/manual/indepth.html

Experiment 4 (Group CIs)

CIs were rated by 60 naïve raters who did not complete any prior experiments. Seven raters was excluded from analyses due to failing an attention check, yielding an analyzed sample of 53 raters (M_{age} = 41.42 years, SD = 10.54; $M_{years\ of\ education}$ = 15.72, SD = 2.30; 24 identifying as female). Of these 53 raters, 45 participants identified as White, 6 as Black, and two as Asian. Of the 53 raters, 46 identified as non-Hispanic.

Instead of evaluating individual Phase 1 participants' CIs, naïve raters evaluated four subgroup CIs from the Phase 1 participants who evaluated agency and four subgroup CIs from the Phase 1 participants who evaluated communality. We generated subgroup CIs to reduce fatigue due to having four evaluations per CI to make. Using subgroup CIs reduces potential concerns about false positives that may arise from using overall group CIs (see Cone et al., 2020 and related work). The raters saw the four CIs of agentic faces and four CIs of communal faces in a random order and were unaware of what prompt yielded the CIs. Each CI was rated on four different scales, presented in a random order, related to gendered facial features. The scales were as follows: "How **masculine (1) or feminine (7)** does this face look?" with response options ranging from 1 (very masculine) to 7 (very feminine), "How **feminine** does this face look?" with response options ranging from 1 (not feminine at all) to 7 (very feminine), "How **masculine** does this face look?" with response options ranging from 1 (not masculine at all) to 7 (very masculine), and "Do you think this face better depicts **a man or a woman**?" with two response options (either a man (-1) or a woman (1)).

Results

All models only included a Trait Type effect. We did not include a Phase 1 Perceiver Gender effect because the subgroup CIs were random selections of male and female Phase 1 participants.

Masculine to Feminine scale

A Trait Type effect, $b = 1.30$, $SE = 0.13$, $t = 10.24$, $p < .001$, 95% CI [1.05, 1.55], showed that people rated agentic faces as being more masculine (Estimate = 3.01, $SE = 0.21$, 95% CI [2.58, 3.45]) than the communal faces (Estimate = 5.61, $SE = 0.19$, 95% CI [5.11, 6.10]).

Masculine scale

A Trait Type effect, $b = -1.28$, $SE = 0.15$, $t = -8.63$, $p < .001$, 95% CI [-1.58, -0.98], showed that people rated agentic faces as being more masculine (Estimate = 5.34, $SE = 0.19$, 95% CI [4.93, 5.76]) than communal faces (Estimate = 2.78, $SE = 0.27$, 95% CI [2.16, 3.40]).

Feminine scale

A Trait Type effect, $b = 1.28$, $SE = 0.14$, $t = 9.37$, $p < .001$, 95% CI [1.01, 1.55], showed that people rated communal faces as being more feminine (Estimate = 5.52, $SE = 0.21$, 95% CI [5.03, 6.00]) than agentic faces (Estimate = 2.96, $SE = 0.23$, 95% CI [2.50, 3.42]).

Gender Category

A Trait Type effect, $b = 2.14$, $SE = 0.25$, $z = 8.47$, $p < .001$, 95% CI [1.75, 3.39], showed that people were more likely categorize a communal face as a woman (Estimate = 2.60, $SE =$ 0.48, 95% CI [1.66, 3.55]) than the agentic face (Estimate = -1.68, $SE = 0.40$, 95% CI [-2.47, -0.90]).

Viewing Conditions

All stimuli were front-facing. Since participants presumably faced their screen, this means that all stimuli were viewed from approximately the same angle. All participants viewed their stimuli on a computer screen and were not able to complete the experiments from mobile devices, which helped to standardize the viewing experience across participants.

We offer an example of likely common participant viewing conditions. To calculate the visual angle of the experiment and stimuli used, we measured the distance of one of the author's eyes to a desktop computer monitor and took the average dimensions of all the face image stimuli used across all studies (in centimeters). The distance from a person's eyes to the screen was 52 cm and the average face image stimuli was 11.11 cm x 13.07 cm. To calculate the visual angle, we used the following visual angle calculator (https://elvers.us/perception/visualAngle/) and calculated a visual angle of 14.3260° (14° 19' 0.56").

Because these experiments were conducted online, we do not know the exact viewing conditions across participants. In part accounting for this limitation, we include a random effect of participant in all models reported in the main text. Including a random effect of participant allows for all reported fixed effects to emerge beyond any variance explained by participant differences, which could include differences in viewing conditions.

APPENDIX C: SUPPPLEMENTAL MATERIAL FOR CHAPTER IV

Experiment 1

Results

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

For each evaluation, we regressed ratings on Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions. The models included random effects structures such that intercepts varied by participant and face identity. Facial Characteristics effects also varied by participant and face identity. See Table 4a-c for all model coefficient information.

Candidate endorsement

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3a). Participants were more likely to endorse women with feminized (Estimate = 4.49, 95% CI [4.26, 4.71]) relative to masculinized (Estimate = 4.25, 95% CI [4.02, 4.48]) characteristics, $b = 0.23$, SE = 0.05, $z = 4.30$, $p = 0.001$. No difference emerged for men with feminized (Estimate = 3.87, 95% CI [3.64, 4.10]) relative to masculinized (Estimate = 3.82, 95% CI [3.59, 4.05]), characteristics, $b = 0.05$, SE = 0.05, $z = 0.92$, $p = 0.80$.

Face Gender and Participant Gender interacted (Figure 4a). Women had higher endorsements of women candidates (Estimate = 4.45, 95% CI [4.17, 4.73]) relative to men candidates (Estimate = 4.00, 95% CI [3.72, 4.28]) candidates, b = 0.45, SE = 0.13, z = 3.45, p = 0.003. This difference was stronger among men ($Estimate_{women}$ = 4.29, SE = 0.13; $Estimate_{men}$ = 4.00, SE = 0.14), b = 0.60, SE = 0.13, z = 4.73, p < .001.

Likelihood of success

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3b). Participants evaluated women with feminized (Estimate = 4.61, 95% CI [4.39, 4.82]) relative to masculinized (Estimate = 4.41, 95% CI [4.20, 4.62]) characteristics as more

likely to do well, b = 0.20, SE = 0.05, z = 4.12, p = 0.002. No difference emerged between endorsements of men candidates with feminized (Estimate = 4.05, 95% CI [3.83, 4.26]) relative to masculinized (Estimate = 3.98, 95% CI [3.78, 4.19]), characteristics, b = 0.06, SE = 0.05, z = 1.30, p = 0.57.

Face Gender and Participant Gender interacted (Figure 4b). Women participants evaluated women (Estimate = 4.59, 95% CI [4.34, 4.85]) relative to men (Estimate = 4.18, 95% CI [3.92, 4.43]) as more likely to do well, b = 0.42, SE = 0.12, z = 3.35, p = 0.004. This difference was stronger among men ($Estimate_{women}$ 4.42, SE = 0.12; $Estimate_{men}$ = 3.85, SE = 0.12), b = 0.57, SE = 0.12, z = 4.69, p < .001.

Liking by other people

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 3c). Participants thought other people would like reporting to women with feminized (Estimate = 4.56, 95% CI [4.34, 4.77]) relative to masculinized (Estimate = 4.28, 95% CI [4.07, 4.49]) characteristics, b = 0.28, SE = 0.06, z = 4.90, p < .001. No difference emerged between expected liking of men with feminized (Estimate = 4.01, 95% CI [3.79, 4.23]) relative to masculinized (Estimate = 4.02, 95% CI [3.81, 4.23]) characteristics, b = -0.01, SE = 0.06, z = -0.19, p = 0.99.

Face Gender and Participant Gender interacted (Figure 4c). Women thought others would like reporting to women (Estimate = 4.50, 95% CI [4.23, 4.76]) more than men (Estimate = 4.21, 95% CI [3.95, 4.47]), b = 0.29, SE = 0.12, z = 2.40, p = 0.077. This difference was stronger among men ($Estimate_{women}$ = 4.34, SE = 0.12; $Estimate_{men}$ = 3.82, SE = 0.12), b = 0.52, SE = 0.12, z = 4.41, p = 0.001.

Experiment 2

Results

Effects of Gender and Sexually Dimorphic Facial Characteristics on Outcomes

We regressed ratings on Job Description (-1 = analytical, 1 = interpersonal), Facial Characteristics (-1 = masculinized, 1 = feminized), Face Gender (-1 = man, 1 = woman), Perceiver Gender (-1 = man, 1 = woman) and their interactions. The models included random effects structures such that intercepts varied by participant and face identity. The random effects structure allowed facial characteristics effects to vary for each participant and for each face identity.

Candidate endorsement

Effects of Face Gender and Facial Characteristics were qualified by their interaction (Figure 6a). Participants were more likely to endorse higher endorse women with feminized (Estimate = 4.44, 95% CI [4.26, 4.63]) relative to masculinized (Estimate = 4.27, 95% CI [4.08, 4.45]), characteristics, b = 0.18, SE = 0.05, z = 3.81, p = 0.001. No difference emerged between endorsements of men with feminized (Estimate = 3.93, 95% CI [3.75, 4.11]) relative to masculinized (Estimate = 3.93, 95% CI [3.74, 4.12]), characteristics, b = -0.003, SE = 0.05, z = -0.06, p = 0.99.

Job Description qualified the interaction between Face Gender and Facial Characteristics (Figure 7). We characterized this interaction by examining Facial Characteristics effects for each Face Gender in each Job Description. For the analytical description, participants were more likely to endorse women candidates within feminized (Estimate = 4.33, 95% CI [4.12, 4.53]) relative to masculinized (Estimate = 4.13, 95% CI [3.91, 4.34]), characteristics, b = 0.20, SE = 0.05, z = 3.71, p = 0.01. No difference emerged between endorsements of men with feminized (Estimate = 3.79, 95% CI [3.58, 3.99]) relative to masculinized (Estimate = 3.85, 95% CI [3.63, 4.06]), characteristics, b = -0.06, SE = 0.05, z = -1.17, p = 0.94.

For the interpersonal description, participants were non-significantly more likely to endorse women candidates with feminized (Estimate = 4.56, 95% CI [4.35, 4.78]) relative to

masculinized (Estimate = 4.40, 95% CI [4.18, 4.63]), characteristics, b = 0.16, SE = 0.06, z = 2.84, p = 0.09. No difference emerged between endorsements of men with feminized (Estimate = 4.07, 95% CI [3.86, 4.28]) relative to masculinized (Estimate = 4.01, 95% CI [3.79, 4.24]), faces, b = 0.06, SE = 0.06, z = 1.03, p = 0.97.

Likelihood of success

Face Gender and Facial Characteristics effects were qualified by their interaction (Figure 6b). Participants evaluated women with feminized (Estimate = 4.57, 95% CI [4.39, 4.74]), relative to masculinized (Estimate = 4.41, 95% CI [4.24, 4.58]), characteristics as more likely to do well, b = 0.16, SE = 0.04, z = 3.95, p < .001. No difference emerged between endorsements of men with feminized (Estimate = 4.08, 95% CI [3.91, 4.26]) relative to masculinized (Estimate = 4.09, 95% CI [3.91, 4.26]), characteristics, b = -0.01, SE = 0.04, z = -0.15, p = 0.99.

Liking by other people

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

Face Gender and Participant Gender interacted (Figure 8). Women thought others would like reporting to women (Estimate = 4.35, 95% CI [4.15, 4.55]) more than men (Estimate = 4.04, 95% CI [3.83, 4.24]), b = 0.31, SE = 0.10, z = 3.03, p = 0.01. This difference was stronger among men ($\text{Estimate}_{\text{women}}$ = 4.44, 95% CI [4.25, 4.63]; $\text{Estimate}_{\text{men}}$ = 4.05, 95% CI [3.86, 4.24]), b = 0.39, SE = 0.10, z = 3.78, p < .001.

Job Description effects on College Student Beliefs

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

Face Gender by Facial Characteristics Contrasts

Experiment 1

Face Gender to Facial Characteristics contrast	*estimate*	*CI*	*p*
Female Feminized - Male Feminized	0.57	0.32 – 0.83	**<0.001**
Female Feminized – Female Masculinized	0.24	0.14 – 0.33	**<0.001**
Female Feminized – Male Masculinized	0.61	0.36 – 0.85	**<0.001**
Male Feminized - Female Masculinized	-0.34	-0.58 – -0.09	0.070
Male Feminized - Male Masculinized	0.03	-0.06 – 0.13	0.984
Female Masculinized - Male Masculinized	0.37	0.14 – 0.60	**0.021**

Experiment 2

Face Gender to Facial Characteristics contrast	*estimate*	*CI*	*p*
Female Feminized - Male Feminized	0.49	0.28 – 0.70	**<0.001**
Female Feminized – Female Masculinized	0.18	0.10 – 0.26	**<0.001**
Female Feminized – Male Masculinized	0.48	0.27 – 0.70	**<0.001**
Male Feminized - Female Masculinized	-0.30	-0.52 – -0.09	0.054
Male Feminized - Male Masculinized	-0.00	-0.08 – 0.08	1.000
Female Masculinized - Male Masculinized	0.30	0.09 – 0.51	0.054