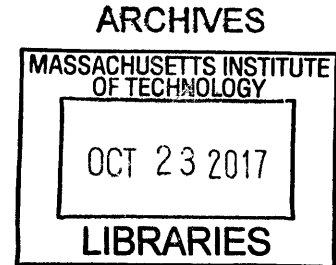# Matching Faces to Expectations:
The Impact of Political and Business Domain Differences in Gendered Expectations on Preferred Appearance for Leaders

By

Hee Jin Yang

B.A. Psychology
Carleton College, 2012

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2017

Signature redacted

Signature of Author:_____

Department of Management

August 4, 2017

Signature redacted

Certified by: _____

John S. Carroll
Gordon Kaufman Professor of Management
Professor of Organization Studies and Engineering Systems
Thesis Supervisor

Signature redacted

Accepted by: _____

_____
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program

1

Matching Faces to Expectations:
The Impact of Political and Business Domain Differences in Gendered Expectations on Preferred
Appearance for Leaders

By

Hee Jin Yang

ABSTRACT

Prior research has demonstrated that feminine appearance is penalized in business contexts.
However, a separate body of research, predominantly centered on electoral outcomes, has put
forth conflicting results demonstrating that feminine appearance can be beneficial. In a series of
experimental studies, I aim to resolve the conflict in the literature by proposing an *expectation-matching* theory. I propose that, although both domains of politics and business are associated
with masculine characteristics, contexts within the political domain are associated with fewer
masculine characteristics than contexts within the business domain and, as a result, perceivers
expect and prefer less masculine facial appearances for political leaders than for business leaders.
The findings of the studies add to the face perception literature and the gender bias literature that
examines the impact of circumstances that make gender and professional roles discrepant in their
impact on consequential outcomes, such as hiring and electoral outcomes.

Keywords: Face perception, appearance, gender, masculinity, domain, politics, business.

Thesis Supervisor: John S. Carroll
Title: Gordon Kaufman Professor of Management

Matching Faces to Expectations:
The Impact of Political and Business Domain Differences in Gendered Expectations on Preferred Appearance for Leaders

"I'm not the woman president of Harvard, I'm the president of Harvard."

- Drew Gilpin Faust, President of Harvard University[1]

Given the large body of work on implicit associations and the strong link between leadership and masculinity, it serves Faust well to redefine herself in her position *sans* gender. Implicit trait theory suggests that people hold archetypes of what a person in a role is like, and so show a preference towards those who appear close to the archetype (Asch, 1946; Cronshaw & Lord, 1987; Cantor & Mischel, 1977). Much to Faust's disadvantage, leadership stereotypes have overwhelmingly been found to be masculine (Boldry, Wood, & Kashy, 2001; for meta-analyses, see Carli, 2001; Eagly & Carli, 2003; and Koenig, Eagly, Mitchell, & Ristikari, 2011). Indeed, individuals gravitate toward leaders who sound strong, appear dominant and agentic- all characteristics which have been associated with men and masculinity (Re, DeBruine, Jones, & Perrett, 2013; Klofstad, Anderson, & Nowicki, 2015). A preference for masculine facial features has been found in hiring for business positions (Sczesny, Spreeman, & Stahlberg, 2006), loan approvals for business purposes (Kuwabara & Thebaud, 2017), and in competitive business contexts (Silberzahn & Menges, 2016). Thus, as conceptions of masculinity and leadership are so closely related, it is sensible to expect any leader wanting to appear competent to distance themselves from associations of femininity.

---

[1] As quoted in Bhayani & Guehenno, 2011.

However, it is not clear whether masculinity is always beneficial for the target in all contexts. Gender role theorists have long outlined the negative consequences for women who display masculine behaviors (Rudman & Glick, 2001; Rudman et al., 2012). Furthermore, facial femininity has been linked with increased attractiveness for both genders, which is related to increased likelihood of winning elections, amongst other outcomes (Perrett et al., 1998; Johnston et al., 2010; Hehman et al., 2014; Carpinella et al., 2015). In sum, the evidence is conflicting and the case for the predominance of facial masculinity is less clear.

One potential factor that could explain the results of the beneficial and/or harmful effects of facial masculinity on consequential outcomes is the *domain* – i.e., the broader level work context – in which the research is based. The context in which a study is set can significantly alter the associations available to the evaluator, and can frame the way in which others are perceived (Higgins, Rholes, & Jones, 1997; Ito et al., 2011). Contexts vary in the relevance of masculine and feminine traits, and thus vary in perceptions of who appears to be a good fit (Garcia-Retamero & Lopez-Zafra, 2006; Johnson et al., 2010). It would then follow that the aforementioned studies, which were primarily based in either business or political contexts, may differ in their characteristics of a good fit. I propose an expectation-match theory to investigate this potential factor as a driver of the conflicting results in the literature on facial femininity.

In this paper, I first establish the differences in the gendered expectations of ideal business leaders compared to ideal political leaders. Here, I expect that, although leaders of both domains will be idealized in terms of mostly masculine trait expectations, leaders in business will be idealized in terms of fewer feminine trait expectations as compared to leaders in politics. Next, I replicate the domain specific preference for facial femininity findings in the literature

4

through an experimental comparison of business and political conditions using one set of stimuli. Finally, I manipulate whether the gendered expectations associated with a domain can influence the hiring decisions to benefit those who are more or less facially masculine.

In summary, I argue that the difference in gendered expectations of domains is a possible explanation in reconciling the division in the literature, where more feminine expectations of a domain would lead to greater preference for those who appear more facially feminine. By contrast, I argue that the domains that are associated with more masculine expectations will benefit more facially masculine candidates.

Facial Masculinity in Social Perceptions

Despite the best of intentions, individuals make trait inferences spontaneously and without awareness (Winter & Uleman, 1984; Bargh & Chartrand, 1999). Faces provide a rich source of information from which inferences can be made (for a review, see Zebrowitz & Montepare, 2008; Willis & Todorov, 2006). Inferences drawn from a single headshot can influence important outcomes such as vote choice (Todorov et al., 2005), mate selection (DeBruine et al., 2006), the length of criminal sentencing (Eberhardt et al., 2004), as well as the likelihood of receiving the death penalty (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Given that most people are unaware of the way that associations can shape beliefs and behaviors, combined with the demonstrable consequential outcomes that they can have, it is clear that it is important to understand the conditions under which inferences are made.

As with any heuristic, inferences, although useful and efficient, are only as good as the associations they are based on. Although inferences can often be accurate, inferences based on

outdated or associations based on pre-existing inequalities can serve to perpetuate incorrect information (Eagly & Karau, 2002; Olivola, Eubanks, Lovelace, 2014; Rule, Krendl, Ivcevic, & Ambady, 2012). In the case of leadership, the history of male dominance in positions of power has cemented the associations between leadership and masculine traits, leaving less room for modern conceptions of female leadership to come to the fore (Ridgeway, 2011). A meta-analysis of 69 studies found strong concurrence on the masculinity of leadership stereotypes (Koenig, et al., 2011). This is not because men are inherently better leaders, as the gender differences between men and women on the qualities important to good leadership are small and benefit both genders (Eagly & Carli, 2007). Rather, it is due to the cultural beliefs that perpetuate the false assumption that men are naturally more adept at being leaders.

The preference for masculinity is not limited to just stereotype content, but has real-world consequences. From a status cue account of gender bias against women in the workplace, markers of femininity (in traditionally masculine environments) serve to reinforce the low status of women, and thus reduce evaluations of competence. This account posits that the greater the association between the target and the low status group (i.e., women), the worse the target's evaluations will be. Sczesny and Stahlberg (2002) demonstrated that the use of masculine perfume by a job candidate led to a stronger certainty in hiring decision than the use of feminine perfume for job candidates of both genders.

The strong associations of leadership and masculinity are reflected in the preferences for facially masculine leaders (Stoker, Garretsen, & Spreeuwers, 2016). Facial masculinity can be measured in various ways: through a global subjective evaluation of how 'masculine' a face appears (Rule & Ambady, 2008; 2009); quantified by software that compares the headshots to

average faces of either gender (Carpinella et al., 2015); digitally manipulated to display the same face with varying levels of masculinity (Silberzahn & Menges, 2016); or can be a measure of how much a face realizes specific masculine facial features, such as a large facial width-to-height ratio (Wong, Ormiston, Haselhuhn, 2011; Stirrat & Perrett, 2010). Both men and women can vary in their facial masculinity (e.g., Republican senator Lindsay Graham is a facially feminine man).

Evidence from the gender bias literature also suggests the importance of facial masculinity in evaluations of competence. In a study using a Goldberg paradigm (i.e., a laboratory based correspondence study (for more information, see Bertrand & Duflo, 2016)), greater attributions of leadership competence were attributed to more masculine faces even when identical resumes were sent with photos of candidates that varied in facial masculinity (Sczesny, Spreeman, & Stahlberg, 2006). Similar results were found outside the lab, using data from a peer-to-peer loan request website, where feminine women asking for projects in traditionally masculine domains were penalized in their actual loan granted amounts, as compared to their more masculine female peers and their feminine peers asking for non-business purposes (Kuwabara & Thebaud, 2017). The authors posited a congruence effect, where the mismatch of the applicant's gender and the loan purpose is heightened when the applicant is especially feminine. The incongruence is reduced when the female borrower appears more masculine and is asking for a masculine endeavor. A similar result was found in business contexts, where people preferred facially masculine male and female leaders when shown digitally manipulated versions of the same face (Silberzahn & Menges, 2016). In a study comparing the perceived competence of business leaders, white men were perceived to be more competent as compared to white

women and black non-babyfaced men, suggesting that there may be some social group combinations that benefit more or less from facial masculinity (Livingston & Pearce, 2009). In sum, the reasons are manifold for why masculine faces would be preferred in traditionally male contexts and their effects on the preference for leaders is clear.

However, other research examining femininity has unearthed contrasting results that leads to a contrast within the literature. Role Congruity theory work shows that agentic (i.e., a masculine trait) women face backlash for breaking the female gender norms governing behavior and are punished in lowered ratings of likability (Rudman & Glick, 2001; Bowles, Babcock, & Lai, 2007; Rudman et al., 2012). Facial femininity could trigger the many benefits afforded to attractive individuals, as both men and women who are facially feminine are seen as more attractive (Perrett et al., 1998; Jaeger, 2011; Johnson et al., 2010). One case in Australia found that the likelihood of winning an election increased as attractiveness increased (King & Leigh, 2009) and another in Finland found that facial attractiveness was associated with a 20% increase in the number of votes for parliamentary candidates (Berggren, Jordahl, & Poutvaara, 2009). Even when controlling for attractiveness, facial femininity in politicians was linked to increased likelihood of winning actual US Senatorial and Gubernatorial elections (Hehman et al., 2014). Similar research using a different measure of facial femininity also found increased likelihood of winning elections for Republican politicians (Carpinella et al., 2015). Thus, it appears that, in certain instances, femininity has proven beneficial in the traditionally masculine environment of politics.

What could be the driver behind finding facial femininity beneficial in some studies but not in others? One potential answer could lie in the varying domain contexts. Context has been

8

shown to play a large part in shaping social perceptions of others (Mitchell, Nosek, Banaji, 2003). A seemingly trivial detail such as the backdrop for a photo was found to attenuate racial discrimination for Black targets in a church context as compared to alternative settings (Bardem, Maddux, Petty, Brewer, 2004). Context has also played a role in the perceptions of charismatic leaders, with social network context shaping the type of leader preferred (Brands, Menges, & Kilduff, 2015). Re-examining the results of Kuwabara and Thebaud (2017), feminine candidates who asked for loans for non-business tasks were not penalized, which suggests that there is a matching of the target to expectations derived from the context. Job level context can also shift perceptions of target suitability, with more facially feminine women being seen as better fits for more feminine job titles (such as secretary or nurse; Johnson, Podratz, Dipboye, & Gibbons, 2010), and with women more likely to be promoted than men in feminine industries such as clothing manufacturing than in masculine industries such as auto manufacturing (Garcia-Retamero & Lopez-Zafra, 2006). The role of expectations within contexts has also been demonstrated to show an impact on perceptions of fit as evidenced through organizational composition (Gorman, 2005). In a study of large law firms in the US, researchers found that the number of gendered characteristics comprising the selection criteria was related to the gender distribution of new hires, with a larger number of stereotypically masculine criteria resulting in a greater proportion of new male hires, and a larger number of stereotypically feminine criteria resulting in a greater proportion of new female hires. As people's associations and expectations across job titles, industries, and domains differ, so does their image of the ideal candidate.

Even though politics and business are both traditionally masculine domains, there seem to be differences in the expectations that individuals hold about them. In a study of warmth and

competence inferences of professions, politicians were rated similarly in terms of warmth, but diverged in their ratings of required competence (Fiske & Dupree, 2014). Politicians were rated as requiring neutral competence, whereas CEOs were rated as requiring high competence. Competence, of the warmth/competence construct, has been traditionally associated with men, suggesting that CEOs may have stronger expectations to display masculine traits than politicians (Fiske, Cuddy, Glick, & Xu, 2002).

If this difference in expectations is meaningful, it would be reasonable to expect similar patterns reflected in the representation rates of women in politics and business. Despite both domains being heavily male dominated, there appears to be a greater representation of women in the highest levels of politics than there are in business. The US. Senate and Gubernatorial offices serves as good example of one of the highest echelons of politics in the country. A similar comparison group in business could be the CEOs of the Forbes 500 companies. When comparing the rates of representation between these groups, we see that 21% of the Senate and 14% of the Governors (including Washington D.C.) are female. In business, we see that only 6.4% of the CEOs in the Forbes 500 are women. Expanding to the Forbes 1000, the percentage drops slightly to 5.9% of CEOs. In sum, the gendered expectations in political and business contexts may differ, in that the business domain may expect greater masculinity traits than in politics.

Through this synthesis of the literature, it is clear that understanding the unconscious associations of appearance to outcomes is important, but that tension exists in whether facial femininity is beneficial or harmful for candidates striving for leadership positions. The difference in the gendered expectations required in business as compared to politics is highlighted as a possible factor that could be driving the contrast in the literature.

Hypotheses and Research Overview

I propose an expectation-match theory to explain the conflicting results of facial femininity preference in the literature. Given the importance of expectations and associations in guiding the archetype for a domain, I propose that the facial features that align most closely to the gendered expectations of a domain will be most preferred. In order to investigate this, I first establish the differences in the gendered expectations of ideal business leaders compared to ideal political leaders. Here, I expect that, although leaders of both domains will be idealized more strongly in masculine trait expectations, leaders in business will be idealized less strongly in feminine trait expectations as compared to leaders in politics.

> *Hypothesis 1*: Ideal leaders in political contexts will be more strongly described by
> feminine traits than ideal leaders in business contexts, across both genders.

If the expectation-match hypothesis is not a valid explanation for the division in the literature, then a difference in the gendered expectations of political and business leaders should not be found. However, even if the hypothesized result were to be found, it is not sufficient to conclude that the differences in facial femininity preferences are due to domain level differences in the gendered expectations.

In order to investigate whether domain level differences are a potential driver for differences in preferred facial femininity, I replicate previous studies in an experimental comparison of business and political conditions using a single set of stimuli.

> *Hypothesis 2*: Facial femininity and domain will interact to predict the likelihood of being
> hired/elected, where facial femininity will be associated with a) an increased likelihood

11

of being elected in a political context, and b) a decreased likelihood of being hired in a business context.

One possible alternative explanation for the differences found in prior research about preferred facial femininity could stem from the use of disparate stimuli. Study 2 controls for this potential confound by using the stimuli from Hehman et al. (2014) and experimentally manipulating whether the targets are introduced as politicians or businesspeople. If the effect was driven solely by stimuli differences, then I should not find any difference in facial femininity preferences when the same faces are presented as being in business vs. politics. However, I find the opposite result in Study 2, where feminine faces were more likely to be preferred (hired or elected) in the political context than in the business context.

The focal outcome variable of whether the perceiver would hire or elect the target was intentionally designed to both closely replicate the designs used in the literature (e.g., Hehman et al., 2014; Sczesny, Spreeman, and Stahlberg, 2011), and to address the research on competence, which has shown that men and women can be equally competent, but still be compensated differently, with women earning less than men (Castilla, 2008). For the sake of brevity, 'hiring' in business and 'being elected' in politics will be referred to as 'hiring' across both contexts when referring to both domains. These results by themselves can only be suggestive as to the potential mechanism that drives the differing results. As a result, I directly manipulate gendered expectations in Study 3.

The main line of reasoning behind the expectation-matching theory is tested in Study 3. If domains possess gendered expectations, and these expectations influence which features are seen as a good fit in potential leaders, then directly manipulating the expectations that perceivers hold

about a domain should lead to changes in the types of candidates that are seen as a good fit. I manipulate the gendered expectations associated with a domain by providing additional information to perceivers emphasizing either the masculine or feminine traits that are purportedly important to the domain. Individuals who were randomly assigned to receive the feminine prompt were expected to prefer candidates with more feminine facial features, as compared to those who received the masculine prompt.

> *Hypothesis 3*: The feminine gendered prompt condition will increase the likelihood of hiring for feminine faced targets, as compared to targets in the masculine gendered prompt condition.

If the difference in gendered expectations drives the difference in facial preferences, then holding gendered expectations constant for each domain should reduce or eliminate any effect of domain. Regardless of the stated domain that the candidates are presented as working in, the perceivers should be influenced more so by the gendered expectations.

> *Hypothesis 3a*: The expectations presented by the Gendered Prompt information will override the effect of domain.

In sum, I argue that the difference in gendered expectations of domains is a possible explanation in reconciling the division in the literature, where more feminine expectations of a domain, such as politics in contrast to business, would lead to greater preference for those who appear more facially feminine. In contrast, domains that are associated with more masculine expectations, such as business in contrast to politics, will benefit more facially masculine candidates.

Study 1

Study 1's goal was to identify whether or not a domain difference in gendered expectations exists between politics and business. Establishing a difference in gendered expectations was the first step in evaluating the possibility of a domain effect of political contexts vs. business contexts as a driver for the mixed results in the facial gender congruity research.

Based on a highly replicated paradigm that assesses the influence of gender stereotypes on expectations of male and female managers through the comparison of ratings of gendered traits, Study 1 compared differences in expectations of leaders in political and business contexts using gendered traits as the dependent measure (Schein, 1973; replicated by McLean & Unter, 2010; Duehr & Bono, 2006; Heilman, Block, Martell, & Simon, 1989; and cross-culturally validated by Schein, Mueller, Lituchy, & Liu, 1996; and Schein & Mueller, 1992). Participants were randomly assigned to visualize a hypothetical ideal leader in either business or politics, and then asked to rate their visualized leader on a list of 20 gendered traits. Given the division in the literature that finds a preference for facially feminine politicians and facially masculine business people, I expected that the mean rating of feminine traits would differ between the visualized political and business leaders.

Hypothesis 1: Ideal leaders in political contexts will be more strongly described by

feminine traits than ideal leaders in business contexts, across both genders.

Thus, participants were randomly assigned to one of the four conditions in this 2 (Domain: Politics or Business) x 2 (Target gender: Male or Female) between-subjects factorial design experiment.

Method

Participants

666 participants based in the U.S. were recruited from and paid through Amazon's Mechanical Turk in exchange for a competitive market rate of $0.35 (Buhrmeister, Kwang, Gosling, 2011). 25 participants were dropped from the analysis for failing at least one of the attention checks, and 20 participants were excluded for incorrectly recalling relevant details of the leader they were asked to visualize. Of the remaining 621 participants, 304 were women. The average age was 34 years (Min = 19, Max = 98), and the majority (74.4%) were employed full-time, with 15.9% unemployed, 6.0% students, and 3.7% retired. For social issues, 21.6% participants identified themselves as being Republican, as compared to 54.4% who identified as Democrat, and 24.0% as Independent. A similar breakdown for political stance on economic issues was found with 30.8% Republican, 45.4% Democrat, and 23.8% Independent.

Procedure

All measures were collected in one sitting via a web survey. After giving consent for participating in the study, participants were randomly assigned to visualize either a Politician (N = 307) or a Businessperson (N = 314) in this balanced design. Half of the participants across all conditions were randomly assigned to visualize a female version of the leader, as specified by female pronouns in the prompt, and the other half a male leader identified with male pronouns.

After they were asked to visualize this ideal leader, participants were asked to rate the leader on a list of 20 gendered adjectives. Demographic information about participant age, gender, employment status, political affiliation, and zipcode was collected. Participants were

then given the manipulation check questions and the last attention check question. All participants were thanked for their time and paid.

Measures

Gendered Expectations scale. In order to assess whether participants had different gendered expectations of leaders in political and business domains, participants were asked to rate how well their visualized leaders were described by a list of gendered adjectives. A list of twenty adjectives were taken from Huddy & Terklidsen (1993) who demonstrated that gendered personality traits influence perceptions of issue competence for male and female politicians. These twenty traits fell into either the "warmth and expressiveness" category, that was deemed feminine, or those representing the "instrumentality" category, which was deemed masculine. The ten feminine adjectives were: Feminine, Warm, Compassionate, Trustworthy, Talkative, Gentle, Family-Oriented, Sensitive, Emotional, Cautious (Cronbach's α = .82). The ten masculine adjectives were: Masculine, Self-Confident, Articulate, Active, Rational, Ambitious, Assertive, Tough, Stern, Aggressive (α = .79). The participants were asked to rate all 20 adjectives on how well each described their visualized leader on a 1-7 Likert scale, with 1 = 'Not at all' and 7 = 'Extremely'). Each participant received a score for the sum total of their ratings on the items on the feminine traits, and another score on the masculine traits; each score ranged from 10 to 70 (e.g., a score of 68 on the femininity scale would indicate high preference for feminine traits).

Manipulation check. To confirm that participants correctly visualized the prompted leader of their condition, participants were asked to recall the pertinent details of the person they had visualized. Participants were asked through multiple choice questions about the prompted leader's gender, age (i.e., adult or child), and domain (i.e., businessperson or politician).

Attention check. Two attention check questions were interspersed among the gendered adjective scales to ensure that participants were responding meaningfully to the question prompts. These attention check questions asked the participants to select option '3' or to choose option '6' and were presented in a randomized order along with the other items in the scale. As noted in the 'participants' section above, we excluded participants ($N = 25$) who incorrectly answered one of these questions, leaving a total sample of 621 participants for the remaining analyses. Since only two attention check questions were used, it is possible that participants who answered randomly may have selected the correct answers by chance alone. By calculating the probability of this occurring, a potential 13 subjects may have passed the attention check through this way in this sample. However, their inclusion in the data would be a conservative influence since it would add noise without any meaningful information.

Results

Manipulation Check

96.8% of participants correctly recalled at the end of the survey the characteristics of the leader whom they were asked to visualize at the beginning of the survey. Of the 20 participants who did not recall correctly, 11 recalled the wrong domain and 9 recalled the wrong gender. As noted in the 'Participants' section above, the 20 participants who failed these manipulation check

questions were removed from the analysis given that failure to answer basic questions about the prompt suggests gross inattention to the study materials.

Differences in Gendered Expectations Due to Domain

Means and standard deviations of the endorsement of feminine and masculine traits for the ideal leaders are presented in Table 1. In order to assess whether the domains of politics and business lead to different gendered expectations of leaders visualized within them, I assessed the mean trait ratings for all of the feminine traits and the mean ratings for all of the masculine traits separately. I then submitted these mean scores to a 2 (Domain: politics or business) X 2 (Target gender: male or female) doubly multivariate analysis of variance (MANOVA) in which the visualized leaders were treated as the unit of analysis. Due to the unequal cell sizes (323 female leaders versus 298 male leaders), I reported Pillai's trace values and used Type I sums of squares to estimate effects, though the same pattern of effects were found when the analysis was modeled using Type III sums of squares.

Results of the MANOVA showed a significant multivariate main effect of domain: $F(2, 616) = 23.3, p < .001, \eta_{partial}^2 = 0.07$. I therefore decomposed the analysis by examining the univariate effects of the masculine trait ratings and the feminine trait ratings with a Bonferroni corrected threshold ($\alpha = 0.025$). Ratings of both masculine traits and feminine traits were significantly effected by the context of the visualized leader: both F's < 9.23, all p's < 0.001. A comparison of the means (see Table 1) within each dependent measure, using a further Bonferroni corrected threshold of $\alpha = 0.013$, revealed the same pattern for each measure that supports Hypothesis 1, that gendered expectations differ between politicians and business

18

leaders. Leaders in the political condition received higher feminine trait ratings than leaders in the business condition, Welch's two-sample $t(604) = 5.06$, $p < 0.001$. Similarly, leaders in the political condition received lower masculine trait ratings than leaders in the business condition, $t(589) = -3.03$, $p = 0.003$.

Table 1

Study 1: Means and standard deviations of gendered trait ratings of ideal business and political leaders

Gendered adjectives	Political leaders						Business leaders					
	Female politicians		Male politicians		Overall		Female politicians		Male politicians		Overall	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Masculine traits	53.07	7.04	52.84	9.46	52.95[a]	8.33	54.10	6.60	55.63	6.89	54.80[a]	6.77
Feminine traits	46.88[c]	8.65	44.54[c]	10.53	45.70[b]	9.69	43.17[d]	8.18	40.63[d]	8.65	42.01[b]	8.48

Note. Both masculine and feminine traits range between 10-70, with higher numbers indicating greater endorsement of the trait.
[a,b,c] Means significantly differ at p < 0.001.
[d] Means significantly differ at p < 0.05.

Differences in Gendered Expectations Due to Target Gender

Results of the MANOVA also showed a significant multivariate main effect of target gender: $F(2, 616) = 7.71$, $p < .001$, $\eta_{partial}^2 = 0.02$. Decomposition with a Bonferroni corrected threshold ($\alpha = 0.025$) revealed that endorsement of feminine traits was significantly affected by the gender of the visualized leader ($F(1, 617) = 11.28$, $p > 0.001$, $\eta_{partial}^2 = 0.04$), but not ratings of masculine traits ($F(1, 617) = 1.16$, $p = 0.28$). Both idealized male and female target leaders

were expected to possess masculine characteristics, but female leaders were also expected to possess more feminine characteristics than their male counterparts. A comparison of the means (see Table 1) within each dependent measure revealed that although the domain effect still persisted, where leaders in politics were endorsed more strongly on feminine traits than leaders in business, female leaders were idealized as being more feminine than their male peers across both domains (Welch's two-sample t-tests: $t_{politics}(294.6) = 2.13$, $p = 0.03$; $t_{business}(297.3) = 2.66$, $p = 0.008$). As expected, there was no interaction of participant gender and domain ($F < 1$), consistent with the literature that shows that gendered expectations are equally known and internalized by men and women.

Discussion for Study 1

The results from Study 1 provide evidence for a domain level difference in gendered expectations, where ideal leaders in political contexts are more strongly endorsed with feminine traits than similar ideal leaders in business contexts. Furthermore, masculine traits were more highly endorsed for business leaders than for political leaders. This difference in the gendered expectations of the two domains provides a possible explanation for the mixed findings in the literature showing that the effects of facial masculinity on consequential outcomes varies across contexts (such as in politics or in business). Hence, if contexts differ in gendered expectations, they may also differ in the facial masculinity that is preferred in these contexts.

Although Study 1 successfully showed through adjective choices that individuals hold differentially gendered expectations of politicians and businesspeople, it did not link these expectations to explicit preference for gendered appearances (i.e., it did not show photos or descriptions of facial features). Although it is highly suggestive that the expectations held of

20

politicians and businesspeople aligned with the results of prior research (Fiske & Dupree, 2014; Hehman et al., 2014; Carpinella et al., 2015; Sczesny, Spreeman, & Stahlberg, 2006; Kuwabara & Thebaud, 2017), it is unclear whether there may be alternative explanations for why facial masculinity is more preferred in some contexts as compared to others: The studies were conducted by different authors, using a variety of different stimuli, and measuring facial masculinity in multiple ways.

Study 2 addresses this limitation by examining preferences for masculine faces in business contexts as compared to political contexts. Although research showing the importance of domain in impacting gendered conceptions of workplace 'fit' have been conducted (Johnson et al., 2010; Garcia-Retamero & Lopez-Zafra, 2006), no research has compared the impact of domain on preference for facial features. Study 2 will address this by directly comparing the hiring outcomes for individuals of varying levels of facial masculinity in the two domains.

Study 2

The aim of the second study was to compare and replicate the findings of gendered appearance preference studies comparing political versus business domains. Previous literature has been divided as to the conditions under which appearing physically masculine is beneficial, with some research finding that facial masculinity leads to greater loan awards for business purposes (Kuwabara & Thebaud, 2017), and an increased chance of being interviewed for a business position (Sczesny, Spreeman, and Stahlberg, 2006); but other research finding the converse to be true, where facial femininity has been linked to increased likelihood of being elected for gubernatorial positions for female politicians (Hehman et al., 2014), and for Republicans (Carpinella et al., 2015). No previous study has directly compared the effect of

21

masculinity and femininity in different domain contexts, and thus leaves open the possibility that facial masculinity may have a differential influence in business as compared to political contexts.

Study 2 sought to directly compare the effect of facial masculinity in likelihood of election / hiring in both the political and business domains. The method was based on a well-established paradigm of hypothetical selection for a position (either CEO or political party leader for business and political domains respectively). Participants were randomly assigned to either a business or political domain, and then asked to rate a series of candidates' faces on their suitability for the position. Based on the results of Study 1, I predicted that the business domain would lead to more facially masculine faces being preferred over facially feminine faces, controlling for relevant facial features such as attractiveness and competence. I further predicted that this preference would disappear or reverse for the political domain.

> *Hypothesis 2*: Facial femininity and domain will interact to predict the likelihood of being hired/elected, where facial femininity will be associated with a) an increased likelihood of being elected in a political context, and b) a decreased likelihood of being hired in a business context.

There was no predicted participant or target gender effect, given the mixed results in the previous literature (Parks-Stamm, Heilman, & Hearns, 2008; Derks, Ellemers, van Laar, & de Groot, 2011; Garcia-Retamero & Lopez-Zafra, 2006; for a review, see Ridgeway, 2001). The study design was between subjects with three factors: 2 (Target gender: male or female) x 2 (Domain: Politics or Business) x 2 (Participant gender: male or female). Results were analyzed using logistic regression with the outcome variable of whether each candidate would be hired/elected for the position (Yes / No).

Method

Stimuli.   Stimuli were drawn from a set of 198 (80 women) photos of American politicians used to demonstrate an impact of gender typicality of facial appearance in electoral outcomes (Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014). Example headshots can be seen in Figure 1 below. The original stimuli were chosen from winner and runner-up politicians in Senate and Gubernatorial electoral contests and were gray-scaled and cropped to show just the face and hair. Further details of the stimuli collection process can be found in the original article. After removing those with low-resolution photos, the remaining female politicians were pre-tested for common facial attributes, such as attractiveness, dominance, competence, warmth, age, and gender typicality (Todorov et al., 2005). Any politician who was recognized by any of the pre-test participants was excluded. Any other politician who had become famous since the original study's publication was also removed (e.g., Jeff Sessions, Carly Fiorina). The set of photos was divided into male faces and female faces, and each set was reduced to sixty faces by random number generator.
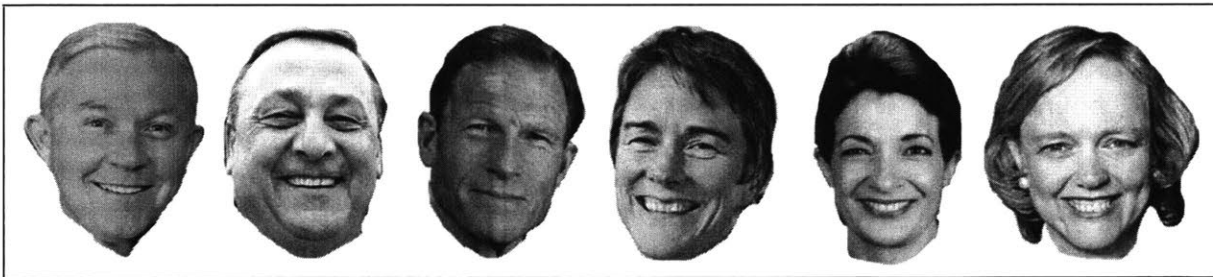

Figure 1. Example photos of stimuli, courtesy of Hehman et al., 2014.

Normed Ratings.   Analogous to Rule & Ambady (2009) and Hehman et al. (2014), I used a separate sample ($N = 309$) to rate the photos on a series of traits to be used as controls. The

23

separate sample was also used to collect information about the facial masculinity of the targets so as to reduce any demand characteristics that may have occurred by calling attention to particular traits while the participants were making hiring/election decisions. The use of software to quantify the measure of facial femininity was not deemed necessary, as previous research had found no differences in the algorithmically derived measures and subjective measures of facial femininity (Carpinella et al., 2015). Raters saw six of the possible 60 headshots of either male or female candidates for each trait using 7-point Likert scales. Along with the focal rating of facial masculinity/femininity, the traits of competence and attractiveness were included to use as control variables given that previous literature has established their importance in social judgments of faces (Heilman & Stopeck, 1985; Budesheim & DePaola, 1994; Todorov et al., 2005).

In addition, participants were asked to rate each of the faces on the top two highest-loading traits from a confirmatory factor analysis conducted on the adjectives from study 1. The two factors of 'masculine' and 'feminine' were extracted with Varimax rotation, with the two most highly-loading traits for the masculine factor being 'self-confident' and 'assertive', and 'warm' and 'compassionate' for the feminine.

The order of the traits was randomized and was presented in blocks grouped by the trait (i.e., participants were asked to rate six photos on one trait, and then another six on another trait, until all the traits were exhausted). Attention checks were at the end of every other block of traits, where participants were asked to select the correct trait they had been asked to rate in the previous block from three potential traits. In order to ensure the quality of the data, participants who failed any one of the attention checks were excluded from the analyses.

24

Participants

Similar to study 1, participants for this study were 218 U.S. based workers who took the survey on Amazon's Mechanical Turk for $.40 as compensation. 18 participants were dropped from the analysis for failing at least one of the attention checks, and a further 33 were removed for having recognized one of the targets in the photos. Of the remaining 167 participants (73 women), two specified 'other' as their gender and were removed due to insufficient sample size to analyze as a third group. The average age was 40 years (Min = 19, Max = 100), and the majority (71.5%) were employed full-time, with 15.8% unemployed, 8.5% students, and 4.2% retired. For social issues, 52.1% of participants identified themselves as Democrat, 24.8% as Independent, and 23.0% as Republican. A similar breakdown for political stance on economic issues was found with 39.4% Democrat, 32.7% Republican, and 24.8% Independent. 70.9% of participants rated themselves the same on both questions, so political affiliation generally overrode policy context.

Procedure

All measures were collected in one sitting via a web survey. After giving consent for participating in the study, participants were randomly assigned to rate a series of either female or male faces (target gender) that were described as either businesspeople or politicians (domain). Each participant rated 14 of the potential 60 faces of the target gender in order to reduce fatigue. Although participants were not timed, they were advised: "Please make your ratings as quickly and accurately as possible." After being presented with a photo, participants were asked whether or not they would elect (in the political condition) or hire (in the business condition) the candidate. The prompt was as follows:

"Would you (elect/hire) this (politician/CEO) to be (in charge of a political party/

running a large company)?"

At the end, participants were asked whether they recognized anyone in the survey and if so, they

were asked to state in a textbox some identifying information. If a participant correctly

recognized any single candidate ($N = 33$), their information was excluded from the analyses as it

may have revealed the true occupation (politician) of the candidates in the headshots.

Political conservatism.   As previous studies have found that preference for facial

femininity is heightened for political conservatives (Carpinella, Hehman, Freeman, & Johnson,

2015; Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014), participants in this study were

asked for their political affiliation. Following the method used in the aforementioned studies,

participants were asked for their zipcode from which their state conservatism could be

calculated. To supplement this measure, participants were asked directly for their political stance

on social and economic issues. As used in previous political science research (Zell & Bernstein,

2014), 7-point Likert scales with the following categories presented from left to right were used:

Liberal Democrat (1), Average Democrat (2), Moderate Democrat (3), Independent (4),

Moderate Republican (5), Average Republican (6), and Conservative Republican (7).

Manipulation check.   To confirm that participants were following instructions and envisioning

the prompted condition, participants were asked to recall the correct domain that the candidates

were presented in. Participants were asked at the end of the survey to select one of three multiple

choice options: 'Politics', 'Business', or the always incorrect decoy option of 'Education'.

Attention check. After participants had filled out demographic information, they were asked two basic questions about the survey to ensure that they were reading the prompts. Participants were asked whether the photos they were evaluating were all either: 1) Women or Men, and; 2) Adults or Children. Failure to answer these questions correctly resulted in removal of their data from the analyses. As noted in the 'participants' section above, 18 participants incorrectly answered one of these questions.

Results

Manipulation check

92.7% of participants correctly recalled at the end of the survey the domain they were rating their targets in (i.e., either politics or business). As noted in the 'Participants' section above, the 16 participants who failed this manipulation check were removed from the analysis given that failure to answer basic questions about the prompt suggested gross inattention to the study materials.

Overview

In this section, I analyze the differences in facial femininity, as quantified by hiring selections, for candidates presented as being in either politics or in business. The dichotomous hiring outcome variable indicates whether or not the participant chose to hire the candidate (0 if not hired; 1 if hired). The binary nature of the hiring outcome meant that a logistic regression (with maximum-likelihood estimation) was the most suitable approach (Czepiel, 2002). Substantially similar results were also found when estimating a probit model with the same controls. Thus, I

regressed hiring outcome on an interaction of domain and facial femininity. The regression models are reported in Table 2.

Hiring outcome, domain, target gender, and participant gender were all dummy-coded variables, with "Not being hired", business, female, and male as the respective reference groups. I initially computed each analysis including control variables for participant political affiliation and employment status. Neither of these variables qualified any of the effects described herein, so they were dropped from the analyses.

These models test the hypotheses that candidates presented as being in the political domain are more likely to be hired as they are more facially feminine, whereas the reverse would be true for candidates presented as being in business domain, where increasing facial femininity would be linked with a decreased likelihood of hiring.

Table 2

Study 2: Logit Models Predicting Likelihood of Hiring

	Model 1	Model 2
Facial		
Femininity	0.288	0.232*
	(.095)	(0.103)
Domain [Ref: Business]		
Politics	0.500	0.641
	(.668)	(0.679)
Target Gender [Ref: Female]	1.659*	0.516
	(0.742)	(0.817)
Interactions		
Domain x Facial Femininity	-.294*	-0.320*
	(0.137)	(0.14)
Domain x Target Gender	-1.202	-1.669
	(1.048)	(1.104)
Facial Femininity x Target Gender	-.491*	-0.076
	(0.204)	(0.232)
Domain x Facial Femininity x Target Gender	0.304	0.417
	(0.286)	(0.308)
Target Facial Rating		
Attractiveness		0.054
		(0.083)
Competence		0.462***
		(0.097)
Feminine traits:		
Compassionate		0.012
		(0.112)
Warm		-0.005
		(0.097)
Masculine traits:		
Self-Confident		0.123
		(0.098)
Assertive		0.182
		(0.111)
Participant Gender [Ref: Male]		0.286***
		(0.084)
Constant	-0.71777	-4.442***
	(0.460)	(0.742)
Observations	2559	2540
Pseudo *R*-Square	0.07	0.1

Note. Standard errors are in parentheses.
* p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed tests)

29

Changes in Likelihood of Hiring due to Domain

I predicted that likelihood of hiring would vary as a function of Facial Femininity, based on domain. More specifically, I hypothesized that appearing more facially masculine would benefit business candidates in their likelihood of being hired, whereas this would disadvantage their political counterparts. Model 1 (see Table 2) examines this key interaction of the independent variables 'Domain' (Categorical condition variable of either 'Politics' or 'Business') and 'Facial Femininity' (Greater endorsement indicates a more feminine face).

As hypothesized, the candidate facial femininity by domain interaction was significant, $b = -.29$, $SE = .14$, $z = -2.1$, $p = .03$. As shown in Figure 2, among business candidates, each unit increase in a candidate's facial femininity (i.e., relatively more feminine) corresponded to a 7.2% decrease in the probability that they would be hired. The most masculine business candidate was 27.8% more likely to be hired relative to the most feminine candidate. Conversely, among political candidates, there was no change in likelihood of hiring with facial femininity. Even moving from the most masculine to the most feminine value of facial femininity only resulted in a 0.7% decrease in the likelihood of hiring.
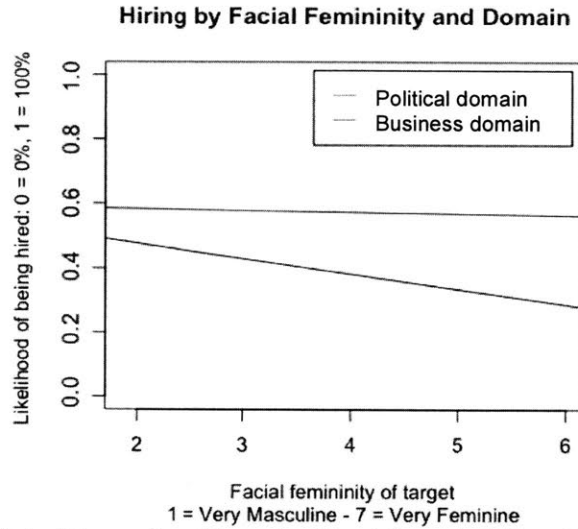
Hiring by Facial Femininity and Domain

Figure 2. Results from Study 2: Plot of the predicted logged odds of the probability of being hired/elected as a function of facial masculinity and domain.

Given the differential target gender means of feminine trait endorsement across domains in Study 1, I analyzed the likelihood of being hired as a function of target gender crossed with domain and facial masculinity. There was no three-way interactive effect of target gender, suggesting that the interaction of domain and facial masculinity applied evenly across candidates of both genders, $b = .30, SE = .29, z = 1.06, p = .29$.

However, there was a two-way interaction of target gender and facial masculinity in predicting likelihood of hire, $b = -.49, SE = .20, z = -2.4, p = .02$ (See Figure 3). Among female targets, each unit increase in candidate's facial femininity (i.e., relatively more feminine) corresponded to a 7.2% increase in the probability that they would be hired/elected, all else being held equal (see Figure 3). Conversely, among male targets, each unit increase in candidate's facial femininity was associated with a 4.8% decrease in the probability that they would be hired/elected. Although this finding makes intuitive sense, the effect disappeared when run in Model 2 with more stringent controls, as discussed below, $b = -.08, SE = .23, p = .74$.

31

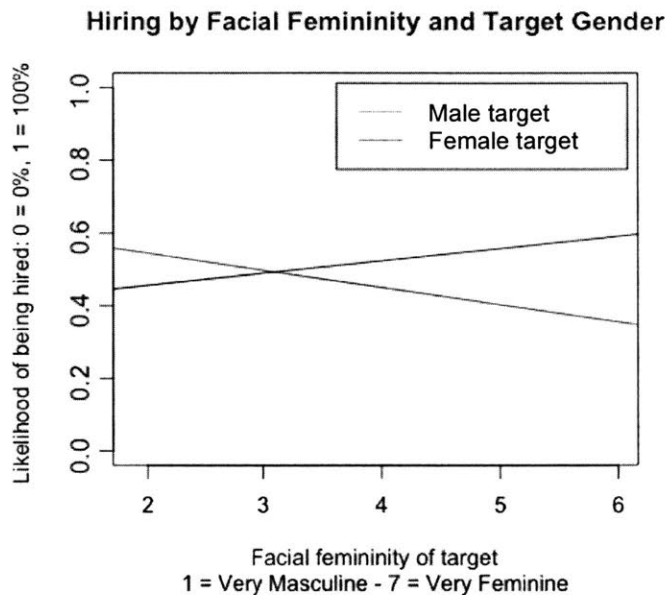## Hiring by Facial Femininity and Target Gender

Figure 3. Results from Study 2: Plot of the predicted logged odds of the probability of being hired/elected as a function of facial masculinity and target gender.

I tested for the robustness of the interaction effect in Model 2 by replicating the original model with the addition of control variables. In order to ensure that the effects were not driven by characteristics of the candidate's other facial features, Model 2 (see Table 2) introduced target rating level controls such as ratings of facial attractiveness, as well as ratings of the competence, warmth, compassion, self-confidence, and assertiveness. Model 2 shows similar results for the key interaction of interest when including these candidate level control variables, with the interaction of domain and facial masculinity on likelihood of hiring remaining significant with a coefficient in the same direction, $b = -.32$, $SE = .14$, $z = -2.29$, $p = .02$.

Although this experiment randomly assigned participants so that approximately equal numbers of both genders would see each condition and therefore participant effects should equalize out, I included participant gender as a predictor variable in Model 2 given the mixed results in the literature surrounding gendered perceptions. Some research has found that men and women enact gender biases differently, whereas others have found no participant gender

32

differences (For a review, see Ridgeway, 2001). In this case, there was a significant effect of participant gender: women were more likely than men to hire or elect the candidate, $b = .29$, $SE = .08$, $z = 3.4$, $p < .0001$, $OR = 1.3$ (See Table 2). Female participants were 57.1% more likely to hire the target as compared to male participants. There was no significant interaction of participant gender and target gender, suggesting that female participants did not preferentially hire one gender over another, but were equally likely to hire candidates of both genders. Again, the focal interaction of domain x facial femininity still remained significant with the addition of the participant gender variable.

Discussion for Study 2

The results from Study 2 replicate the conflicting results in the research literature in a single study through the experimentally manipulated domain variable. The results support the hypothesis that facial femininity is negatively associated with reduced likelihood of hiring in business contexts, as compared to political contexts, whereas facial femininity is not as strongly associated with a decreased likelihood in hiring. Whereas femininely faced business targets were less likely to be hired for the position, similarly feminine political candidates were not as severely punished in their hiring likelihood. Although the results did not reflect a positive association between feminine facial features and likelihood of hiring for candidates in the political domain, there is a significant difference in the effect of facial femininity for the two domains in the hypothesized direction, using identical photos. This finding is notable because until now the domains of politics and business have not been compared side by side within one study.

However, Study 2 has some limitations. One potential limitation could be due to the wording of the focal hiring question for the political domain. Although the political and business domain conditions were created to be as equivalent as possible, certain wording differences were necessary in order for the conditions to make sense. Whereas the focal question in the business condition read "Would you hire this CEO to be in charge of a large company?", the focal question in the political condition read "Would you elect this politician to be in charge of a political party?". Some may be concerned that leading a political party is a much rarer and unique position that may be seen as more prestigious or requiring more stringent qualifications. Although there is not a significant main effect of domain in the model, a look at the mean rates of hire suggest that participants were more likely to hire candidates in the political domain rather than the business domain, (57.2% compared to 36.8%), $\chi^2(1, N = 2559) = 107.4, p < 0.001$. Even though the data suggest that participants are more selective in hiring for the business context than the political context, I amended the focal question in Study 3 to better reflect the more typical political election that would be seen as more equivalent to the position of CEO of a large company.

Lastly and most importantly, however, Study 2 cannot definitively conclude the mechanism for why the differential preference in facial femininity occurs. As no mechanism is experimentally manipulated, the difference between the two domains may exist due to some unspecified or untested feature differentiating politics and business. This limitation is addressed in Study 3, where the results from Study 1 and Study 2 are brought together to experimentally manipulate the role of gendered expectations of the domains in driving the preferential selection of faces.

Study 3

Whereas Study 1 identified differences in gendered expectations of politicians and businesspeople, Study 2 showed differences in preferred facial masculinity of politicians and businesspeople in a hypothetical hiring/election decision. The aim of Study 3 was to link the gendered expectations in different work domains to preferences in facial masculinity of candidates in order to test the hypothesis that domain affects the relationship between gendered facial appearance and preference for hiring through its gendered expectations. I manipulated the desired gendered expectations by presenting a prompt emphasizing the importance of either feminine or masculine characteristics for success in either the business or political domain. Individuals who were randomly assigned to receive the feminine prompt were expected to prefer candidates with more feminine facial features, as compared to those who received the masculine prompt.

> *Hypothesis 3*: The Feminine Gendered Prompt condition will increase the likelihood of hiring for feminine faced targets, as compared to targets in the Masculine Gendered Prompt condition.
>
> *Hypothesis 3a*: The expectations presented by the Gendered Prompt information will override the effect of domain.

This was tested with a logistic regression, with whether or not the candidate should be hired/elected as the dependent variable.

35

Method

Stimuli.　　　　The same photographic stimuli from Study 2 were used in this experiment, with the only change being removing any headshots that had been recognized by at least one participant in Study 2, leaving a total of 47 male and 49 female stimuli photographs.

Normed Ratings.　　　In order to control for the effect of important facial attributes, I used the normed ratings from the independent sample in Study 2.

Participants

208 U.S. based mTurkers took the survey for $.35 as compensation. Nine participants were removed for correctly recognizing one of the targets in the stimuli photos and seven participants were dropped from the analysis for failing at least one of the attention checks. One participant was removed due to a comment indicating an intention not to cooperate with the survey due to their ideological stance. A further 30 participants were excluded for failing the manipulation check. Of the remaining 161 participants, 92 were women. The average age was 34 years (*Min* = 19, *Max* = 73). 77.6% were employed full-time, with 13.7% unemployed, 5.6% students, and 3.1% retired. 18.9% of participants identified themselves as Republican, 51.8% as Democrat, 26.2% as Independent, and 3.0% as Other.

Procedure

Participants were randomly assigned to receive a prompt that emphasized either feminine or masculine characteristics, as established by Atwater, Brett, Waldman, DiMare, & Hayden (2004). The masculine prompt read:

> *Past research has shown that successful politicians are good at **allocating resources, delegating, disciplining, punishing, strategic decision-making**, and **problem solving**.*

The feminine prompt followed the same format, but had the characteristics replaced with those associated with femininity:

> *Past research has shown that successful politicians are good at **planning** and **organizing, developing** and **mentoring, supporting, motivating** and **inspiring**, providing **corrective feedback, communicating** and **informing**.*

In order to ensure that these prompts had the intended effect of shifting participants' gendered expectations of the domains, the prompts were pilot tested with an independent sample similar to the study participants ($N = 191$). In order to gauge whether the prompts shifted the gendered expectations associated with the domains (albeit temporarily), raters either read the masculine or feminine prompt and then were asked to select from a list of characteristics that best described the traits of a successful leader in either business or politics. The traits that participants could select from were the same traits as used in Study 1, ten masculine and ten feminine traits, each

rated on a 1-7 scale for how well they described a successful leader. The masculine and feminine traits were summed separately to produce scores that could range from 10-70. The prompts successfully shifted the gendered expectations for feminine traits in both domains, with those who received the feminine prompt endorsing more feminine traits ($M = 45.8$, $SD = 8.1$) than those who received the masculine prompt ($M = 38.3$, $SD = 11.5$), $F(1, 187) = 27.6$, $p < 0.001$. However, the corresponding difference for masculine traits was only marginally significant, with those who received the masculine prompt ($M = 56.5$, $SD = 8.7$) demonstrating somewhat greater endorsement of masculine traits compared to those who received the feminine prompt ($M = 54.3$, $SD = 7.1$), $F(1, 187) = 3.5$, $p = 0.06$. Given the results of Study 1, where masculine traits were relatively immune to change, and also due to the positive effect of feminine traits, these prompts were used in Study 3 to manipulate the gendered expectations across domains. There were no interaction between the domain and prompt for either the masculine traits ($p = .79$) or the feminine traits ($p = .63$).

To ensure that participants in Study 3 read and understood the prompts, participants were asked two comprehension check questions directly after the prompts. These questions asked participants to correctly identify which characteristics were indicative of success in the work domain they were assigned to. Although participants who failed this comprehension check were allowed to progress through the survey, their data was excluded from the analyses.

After reading the prompt and answering the questions, participants followed the same survey procedure as in Study 2, rating a series of either female or male faces (target gender) that were described as either businesspeople or politicians (domain). Each participant rated 20 of the

potential 47 male or 49 female faces in order to reduce fatigue. The focal question of whether the participant wanted to hire/elect the candidate was modified slightly in order to address a shortcoming in the Study 2 design. The focal question in the political condition of Study 2 of "being in charge of a political party" was changed in Study 3 to read: *"Would you elect this politician to serve as a senator?"* This modification aligned the focal question to correspond better to the literature linking gendered appearance to electoral outcomes, where studies have examined likelihood of being elected as a senator or governor as opposed to leading a political party. It also had the additional benefit of making the political focal task more comparable to running a large company and making the question more in line with what voters are asked to do.

Manipulation check. After reviewing the candidates' photos, participants were asked to recall the characteristics related to success in their assigned domain presented to them at the beginning of the survey. All the feminine and masculine characteristics were presented in a multiple-choice list, in which participants were asked to select the six that they had been presented. Any incorrect responses were subtracted from the sum of the total correct responses. The scores could range from -6 (when the participant chose all of the wrong characteristics and none of the correct ones) to 6 (chose all correct characteristics and none of the wrong ones). Any participants scoring less than 3 were excluded from the analyses.

Stimuli Recognition Check. Participants were asked whether they had recognized any of the targets, and to identify anyone they recognized. Nine participants who correctly recognized anyone in the survey were excluded from the analyses. Four participants incorrectly guessed a

candidate's identity (e.g., "Jim Barnett former Basketball Player"), and their information was kept.

Attention check. As in Studies 1 and 2, participants were asked at the very end of the survey basic questions to ensure that they had read the prompts. Participants were asked three multiple choice questions about the targets' gender, age (i.e., adult or child), and domain (i.e., Business, Politics, or the decoy option of Education). Participants only saw targets of one gender, age (all targets in the study were adults), and domain. Seven participants incorrectly answered one of these and were removed from the analyses; Five recalled the wrong domain and two recalled the incorrect gender.

Results

Manipulation check

85.8% of participants scored at least 3 or higher in the manipulation check (ranging from -6 to 6). The median manipulation check score was 4, suggesting that most participants understood and correctly recalled the gendered expectation prompts given at the beginning of the survey. The 30 participants who failed this manipulation check were removed from the analysis given that the low scores reflected inattention to the study materials, as noted in the 'Participants' section above.

In this section, I analyze the differences in gendered facial appearance preference in hiring of those exposed to feminine and masculine expectation prompts across domains. As in Study 2, the dichotomous outcome variable indicates whether or not the participant chose to hire/elect the candidate (0 if not hired; 1 if hired). Again, I used a logistic regression (with maximum-likelihood estimation) in Model 1 and introduce controls in Model 2; substantively similar results were found when estimating a probit model with the same controls. I regressed hiring outcome on a three-way interaction of gendered expectation prompt, domain, and facial femininity. The regression models are reported in Table 3.

Hiring outcome, gendered expectation prompt, domain, target gender, and participant gender were all dummy-coded variables, with "Not being hired", feminine expectation prompt, business, female, and male as the respective reference groups.

These models test the hypotheses that the gendered expectations associated with a domain influence the preference of gendered facial features in hiring, above and beyond the effect of the domains themselves. Across both domains, participants who receive the feminine gendered expectation prompt are hypothesized to prefer candidates who are more facially feminine, as compared to those who are more facially masculine. The reverse is hypothesized for participants who receive the masculine gendered expectation prompt, where there would be a preference for hiring candidates who are facially masculine compared to their facially feminine counterparts, across both domains.

Table 3

Study 3. Logit Models Predicting Likelihood of Hiring

	Model 1		Model 2	
Facial Femininity	0.353***	(0.063)	-0.028	(0.084)
Gendered Expectation Prompt [Ref: Feminine]				
Masculine Expectations	0.585	(0.387)	0.639	(0.39)
Domain [Ref: Business]	0.645	(0.388)	0.528	(0.392)
Politics				
Interactions				
Prompt x Facial Femininity	-0.186*	(0.097)	-0.197*	(0.097)
Prompt x Domain	-0.736	(0.562)	-0.691	(0.567)
Facial Femininity x Domain	-0.068	(0.096)	-0.04	(0.097)
Prompt x Domain x Facial Femininity	0.153	(0.138)	0.126	(0.14)
Participant Gender [Ref: Male]			0.429***	(0.074)
Target Gender [Ref: Female]			-0.989***	(0.134)
Target Facial Rating				
Attractiveness			0.309***	(0.073)
Competent			0.054	(0.082)
Feminine traits:				
Compassionate			-0.321**	(0.105)
Warm			-0.006	(0.078)
Masculine traits:				
Self-Confident			0.135	(0.096)
Assertive			-0.178*	(0.082)
Constant	-1.382***	(0.258)	0.708	(0.656)
Observations	3220		3220	
Pseudo R-Square	0.04		0.09	

Note. Standard errors are in parentheses.
* p < 0.05; ** p < 0.01; *** p < 0.001 (two-tailed tests)

42

Changes in Likelihood of Hiring Due to Gendered Expectation Prompt

I predicted that likelihood of hiring would vary as a function of the gendered expectations prompt information received and facial femininity of the candidate. More specifically, I hypothesized that appearing more facially feminine would benefit candidates being hired by evaluators who had received the feminine prompt across both domains, whereas the same facial features would serve as a disadvantage when being evaluated by hirers who received the masculine prompt. Model 1 (see Table 3) is based on this key interaction of the independent variables 'Prompt' (Categorical condition variable of either 'Feminine' or 'Masculine') and 'Facial Femininity' (Greater endorsement indicates a more feminine face).

As hypothesized, the candidate facial femininity by domain interaction was significant, $b = -.19$, $SE = .10$, $z = -1.9$, $p = .05$. Among those who read the Feminine expectations prompt, each unit increase in a candidate's facial femininity (i.e., relatively more feminine) corresponded to an 8.2% increase in the probability that they would be hired (see Figure 3). Using the predicted coefficients, the most feminine business candidate was 34.4% more likely to be hired relative to the most masculine candidate of this sample. Conversely, among those given the masculine expectations prompt, there was only a 4% increase in the likelihood of hiring with each unit increase in candidate's facial femininity (i.e., relatively more feminine). Although increases in facial femininity were related to increased likelihood of hiring by evaluators who received the Masculine expectations prompt, the effect was less than half as strong compared to the Feminine expectations prompt condition.
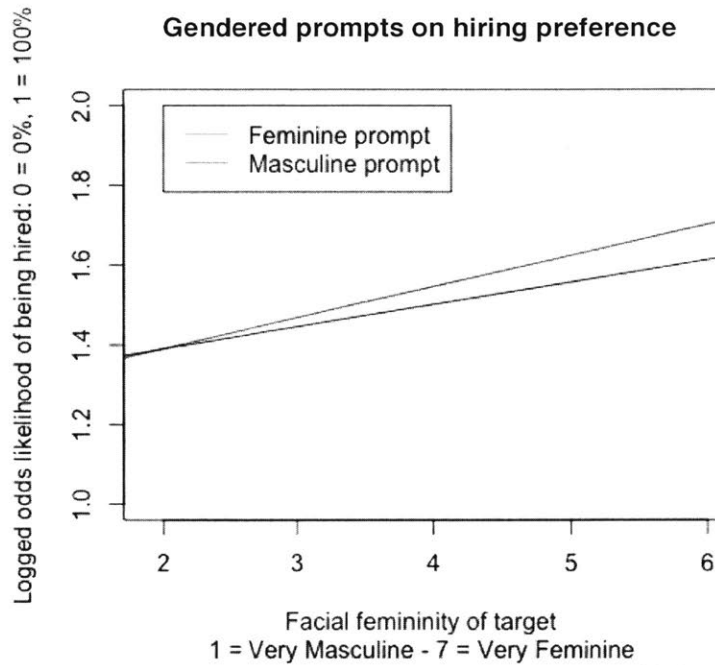
Figure 3. Results from Study 3: Plot of the predicted logged odds of the probability of being hired/elected as a function of facial masculinity and gendered expectation prompt.

Also as hypothesized, there was no main effect nor interactive effect of domain on the likelihood of hiring, providing further suggestive evidence that the gendered expectations associated with domains could be the influencing factor in the preference for more or less masculine faces, all p's > 0.09.

However, there was a main effect of facial femininity on likelihood of hire that worked in preference of more feminine faces, $b = .35$, $SE = .06$, $z = 5.6$, $p < .001$, $OR = 1.4$ (See Table 3). This finding disappeared once the model was re-run with more stringent controls, as discussed below, possibly indicating that variance explained by some of the control variables may have been provisionally explained by a main effect of facial femininity, $b = -.03$, $SE = .08$, $p = .74$.

I tested for the robustness of the focal interaction effect of Gendered Expectation prompt and facial femininity on likelihood of hiring in Model 2 by repeating the original model with the

44

addition of control variables. In order to ensure that the effects were not driven by other facial features, Model 2 (see Table 3) introduced controls such as target gender, ratings of facial attractiveness, and ratings of competence, warmth, compassion, self-confidence, and assertiveness. Model 2 shows substantively similar results when including these candidate level control variables, with the focal interaction of domain and facial masculinity on likelihood of hiring remaining significant, $b = -.20, SE = .10, z = -2.02, p = .04$.

Of the control variables, there were three main effects of target facial characteristics on the likelihood of hiring. In line with the vast literature showing the positive benefits of attractiveness (for a review, see Van Leeuwen & Macrae, 2004), facial attractiveness of the targets in this sample was related to an increased likelihood or hiring, $b = .31, SE = .07, z = 4.2, p < .001$. In contrast, ratings of compassion and assertiveness based on the target's face were related to a decreased likelihood of hiring, $b_{compassion} = -.32, SE = .11, z = -3.1, p < .01$ and $b_{assertive} = -.18, SE = .08, z = -2.2, p < .05$ (See Table 3).

Given the participant gender effects of Study 2, I included participant gender as a predictor variable in Model 2. Women were significantly more likely than men to hire the candidate overall, $b = .43, SE = .07, z = 5.8, p < .0001, OR = 1.4$ (See Table 3). Female participants were 57.7% more likely to hire the target as compared to male participants. This result is consistent with the results of Study 2, where female participants were 57.1% more likely to hire targets overall (See Table 2).

In contrast to Study 2, there was a significant effect of target gender on likelihood of hire in Study 3. Female targets were more likely to be hired than male targets, $b = -1.0, SE = .13, z = -7.4, p < .0001, OR = 0.37$ (See Table 3). Male targets were 27% less likely to be hired than

45

female targets. This finding is surprising given that there was no impact of target gender in Study 2, which is discussed further below.

Discussion for Study 3

The results from Study 3 support the hypothesis that the gendered expectations associated with a domain influence the preference of gendered facial features in hiring, above and beyond the effect of the domains themselves. Despite replicating the same domain information given in Study 2, the gendered expectation prompts removed the effect of domain. This provides strong, albeit indirect, support for the hypothesis that the gendered expectations of domains influence the preference for gendered facial features in hiring.

Support was also found for the hypotheses surrounding the influence of the specific feminine and masculine prompts on the likelihood of hire. The hypothesized effect of facial femininity on increased likelihood of hiring was found when the feminine expectations prompt was given. However, although there was not a complete reversal in the direction of the relationship between facial femininity and likelihood of hiring in the masculine expectation prompt condition, the influence was reduced by over 50%.

Given that the masculine expectation prompt held only a modest influence in changing gendered expectations in the pilot test, the lack of a strong positive relationship in Study 3 between facial masculinity and likelihood of hiring in the masculine prompt condition may be due to the lack of strength in the manipulation. As seen in Study 1, masculine traits were prized highly in both political and business domains, and the difference between the two domains was driven mainly by the difference in importance of feminine traits. It may be possible that masculine traits are always highly valued and are resistant to alternate information, making the

task of manipulating their importance difficult. Another possibility could be that masculine traits

may be more strongly held, whereas the conceptions of ideal feminine traits are more susceptible

to change. Given research showing the change in feminine stereotypes over the past decades, the

latter may be likely (Heilman, Block, Martell, & Simon, 1989).

As the masculine prompt did not significantly change expectations, the masculine

expectations condition could be reconsidered as a baseline in which to compare the effect of the

feminine prompt against. The divergent effects of facial femininity on the likelihood of hiring is

indicative of the differential effect the gendered expectation prompts had, where feminine faces

were preferred at a greater rate over masculine ones in the feminine prompt condition as

compared to the masculine prompt condition, regardless of domain.

Taken together, these results add to the literature in presenting the gendered expectations

of a domain as a potential factor that could explain the divided literature on the beneficial effects

of gendered appearances. The theoretical and practical implications, as well as limitations of this

set of studies, will be discussed in the general discussion below.

General Discussion

Across three studies, I found cumulative evidence to support the expectation-match theory:

Perceivers match the salient gendered expectations of the situation to the facial femininity of

targets when making judgments of leadership fit. I first established the differences in the

gendered expectations of ideal business leaders compared to ideal political leaders. As

hypothesized, leaders of both domains were idealized more strongly in masculine trait

expectations, but leaders in business were idealized less strongly in feminine trait expectations as

compared to leaders in politics. I then investigated whether domain level differences were a

47

potential driver for differences in preferred facial femininity by replicating previous studies in

the literature in an experimental comparison of business and political conditions using a single

set of stimuli. As hypothesized, feminine faces were more likely to be preferred (elected/hired) in

the political context than in the business context. The third study then tested the mechanism

behind the expectation-matching theory. If domains possess gendered expectations, and these

expectations influence which features are seen as a good fit in potential leaders, then directly

manipulating the expectations that perceivers hold about a domain should lead to changes in the

types of candidates that are seen as a good fit. By presenting additional information that

emphasized either the masculine or feminine traits, I manipulated the gendered expectations that

participants had salient to them about their domain. As hypothesized, individuals who were

randomly assigned to receive the feminine prompt preferred candidates with more feminine

facial features, as compared to those who received the masculine prompt. Although a strong

positive preference for facial masculinity was not found in the masculine prompt condition, there

was still a difference in the effect of facial femininity preferred due to the prompt content. The

lack of effect for the masculine prompt may be due to the closely held beliefs surrounding the

value of masculine traits, and reflect the relative malleability of feminine traits in shifting value.

Lastly, as hypothesized, the gendered expectation prompts appeared to mediate, albeit indirectly,

the relationship between domain and preference for facial femininity, where holding gendered

expectations constant for each domain eliminated any effect of domain. Regardless of the stated

domain that the candidates were presented as working in, the perceivers were influenced by the

gendered expectations as opposed to the domains themselves. In sum, the difference in gendered

expectations of domains appears to be a possible explanation in reconciling the division in the

literature, where more feminine expectations of a domain, such as politics in contrast to business, lead to greater preference for those who appear more facially feminine.

A few surprising findings of the studies should be considered. Given the voluminous research showing the lack of fit between the associations held of women and traditional forms leadership (for a review, see Eagly & Carli, 2007), the preference for female targets found in Study 3 is noteworthy. One possibility could be that cultural beliefs around suitability for women in positions of power may have changed in the recent years. However, that is less likely given the absence of a preference for female targets in Study 2. A more likely culprit may have been due to the similarity of average facial masculinity ratings for male and female targets in this paper, due to the design of the survey collecting the norming ratings. The independent group of raters judged the facial masculinity of the targets of one gender only. As such, participants may have been comparing the facial masculinity of the targets implicitly against targets of the same gender, instead of against people in general. Research on the 'Shifting Standards' effect has found that ratings of male and female targets do not differ when compared against others of the same gender, but differences are revealed when objective points of comparison are introduced (Biernat & Fuegen, 2001). The design decision to ask raters, as well as the participants in the broader studies, to evaluate targets of only one gender was a deliberate choice to avoid rater concerns of appearing sexist and preferentially rating women more favorably than men. When participants are asked to compare candidates of two genders with limited information, participants become sensitive to salient social category differences like race and sex (Koenig et al., 2011). As all the candidates were white, the most salient category difference would have been gender. On the converse, the salience of gender can also exacerbate gender stereotypes,

49

leading to increased preference for male targets (Boldry, Wood, & Kashy, 2001). As a result, the design decisions were made in order to prevent gender from becoming a very obvious differentiating factor that may have influenced participant ratings.

Another unhypothesized finding was the lack of strength in the masculine expectation prompt condition in increasing the likelihood of preference for facially masculine targets. As the prompt only generated marginally significant effects in the pilot, with only a very slight difference between the mean scores of traits, the lack of effect in Study 3 of the masculine prompt condition is most likely due to the weakness of the prompt content. Given that the prompt content was based on previous literature on the gendered nature of professional traits, it is surprising that a stronger effect was not found. One potential explanation could be that, given the long history and predominance of men in leadership positions throughout history, the importance of masculine traits is more resistant to change, as compared to feminine traits. Whereas the stereotypes surrounding women and female leadership have changed dramatically over time, the stereotypes surrounding men and male leadership have not (Heilman, Block, Martell, & Simon, 1989). Furthermore, women need to be seen to exhibit excellence in both masculine and feminine traits, what Bem describes as being 'androgrynous' (1974), and thus the feminine traits that are seen as important in a manipulation may be more sensitive to reaffirmation through prompt information. Indeed, leadership prototypes for men require strength, but for women require both strength and sensitivity (Johnson, Murphy, Zewdie, & Reichard, 2008). In order for women to fill their gender role, women must appear feminine, but in order to fulfill their professional role, they must also appear masculine (Heilman & Saruwatari, 1979; Heilman & Stopeck, 1985, Johnson, Podratz, Dipboye, & Gibbons, 2010;

Kuwabara & Thebaud, 2017). In contrast, men are not required to appear masculine in order to appear competent (Mattes et al., 2010). In summary, the importance of masculine traits may not be malleable and may be resistant to manipulation through prompt information. Future research could investigate whether the importance of masculine traits can be minimized or altered.

Theoretical Contributions and Practical Implications

Until now, the vast majority of the literature has overlooked the impact of domain in influencing the preference for certain facial features. The examination of the impact of domain that this paper presents is important given that the results of these studies can help situate the findings of previous work done on preferences for facial femininity and masculinity. Although the importance of context and domain has emerged in previous research, no research to date has examined the impact of domain on the preferences for facial features. The results of the studies presented here suggest that any face research should account for the context that the experimental study is presented in. These results also suggest the broad array of implicit associations perceivers can draw from faces and their far-reaching impact on decision making. Inferences drawn from faces can help predict fit at a varying range of domains, and could possibly be extended to organizations, specific positions, or whatever category comprises gendered expectations. Furthermore, the significant impact that gendered expectations held on influencing preference for hiring adds to the gender bias literature in presenting another avenue by which stereotypes and associations related to gender can affect perceptions of fit in professional life.

The practical implications for the results presented in this paper directly affect the strategies that individuals can undertake to increase perceived fit in specific domains. Whereas it

is a vast undertaking to change one's physical presentation of gender completely, it is easier to modify one's appearance to appear more or less masculine or feminine. Individuals can make an informed decision in their hairstyle, accessory choice, or make-up use depending on their context. When choosing between photos to attach for a job talk presentation or for campaign posters, individuals may find some guidance for their choice depending on the gendered expectations that their task may hold.

However, the intention of this research is not to dictate the ways in which individuals should change in order to better fit traditional gender roles, but rather to inform individuals to strategically navigate the effects of the implicit associations they may encounter. The onus for counteracting the forces of gender norms on perceptions of fit should not rest predominantly on the shoulders of individuals, and the results of this research could also be applied by organizations interested in understanding ways to create inclusive environments for their entire community. By emphasizing the right set of traits and associations as valued, the organization may help attract or retain individuals of previously underrepresented populations. Future research should be conducted before the application of these results into an intervention style program.

<div align="center">Limitations</div>

Despite their contribution to the literature, the studies presented in this paper are not without their limitations. Although Study 3 strongly suggested evidence of the hypothesized relationship, there was no formal test of mediation done to investigate the mediating effect of gendered expectations on the relationship of domain and preferences for facial femininity. This was due to a deliberate design decision where it was deemed more beneficial to reduce the potential demand

effects caused by drawing attention to the concept of facial masculinity while participants were making their hiring decisions. If participants were asked to rate the facial masculinity of the target before making their hiring decision, facial masculinity may become a salient factor that could potentially become susceptible to demand effects. When perceivers' attention is not drawn to target facial masculinity, they remain unaware of their use of it as a guide to make decisions (Sczesny & Kühnen, 2004). Making implicit expectations explicit can remove the impact that the implicit effect may have had (Schwartz & Clore, 1983). As a result, in order to maintain the implicit nature of the associations drawn from the stimuli's facial masculinity and the gendered expectations, it was necessary to use an independent norming group to ascertain facial traits.

Critics might also point to the use of naturalistic stimuli as a shortcoming of Studies 2 and 3. Instead of presenting a selection of 5-7 versions of one face that was digitally altered in its facial masculinity, Studies 2 and 3 presented a large range of target faces that varied naturalistically in facial masculinity as well as other traits. However, even digitally manipulated versions of varying facial masculinity of one face also varies in the other facial characteristics, such as in attractiveness and competence (Perrett et al., 1998). The disadvantage of presenting digitally altered versions of a face and asking participants to select the best face is that it draws attention to the focal variable of facial masculinity as well as presents a relatively unnatural situation from which to choose. Future studies could be based on a different assessment of benefits and shortcomings.

Lastly, a critical limitation of these studies is due to the use of all White target stimuli. Previous research has shown that social categories do not form a simple additive effect of race and gender, but can have unique interactive effects (Rosette & Livingston, 2012; Livingston &

53

Pearce, 2009; Pratto & Espinosa, 2001; Crenshaw, 1991). Although these results are important findings that still apply in the consideration of White targets, future research should extend the findings with targets of different races.

Future Directions

Since these studies examined only facial femininity, future research could extend the scope of fit characteristics to include other aspects of gendered appearance such as attire. A replication of these experiments could easily ascertain whether feminine vs. masculine dress impacts the perceived fit of targets in varying domains. Another avenue for future research concerns the generalizability of the influence of facial femininity to broader feminine traits and behaviors, as evidenced for example through quotes, personality traits, or resume line-items. It is worth investigating whether the effect is limited to gendered visual features, or whether behavioral and non-appearance traits can also trigger the associations between domains and perceptions of fit.

Another line of future research concerns the automaticity of updated evaluations. Although it is evident that there is an implicit association between the gendered expectations inherent to the domains with the preferred target faces, it is less clear whether the match is as implicit once the additional prompt information is provided. An evaluation of whether participants consciously correct for their preferences using the prompt information or whether participants associate unconsciously may be interesting. Future research could use a cognitive load task to disturb the capacity for conscious correction, or alternatively use an Implicit Association Test to compare the response times for pairing "Feminine" and "Masculine" with "Good" and "Bad" before and after receiving gendered expectation prompt information. Multiple

ideas for future research are possible in examining the boundaries of the matching effect of

gendered expectations to gendered facial appearances. There are many opportunities to extend

this research into other aspects of the fit between expectations for appearance and work roles.

Works Cited

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American psychologist, 54*(7), 462.

Bem, S. L., & Lewis, S. A. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of personality and Social Psychology, 31*(4), 634.

Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics, 94*(1), 8-15.

Bhayani, P. D., & Guehenno, C. M. (2007, February 11). Faust Confirmed as 28th President. *The Harvard Crimson.* Retrieved from http://www.thecrimson.com

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707-724.

Boldry, J., Wood, W., & Kashy, D. A. (2001). Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues, 57*(4), 689-705.

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and human decision Processes, 103*(1), 84-103.

Brands, R. A., Menges, J. I., & Kilduff, M. (2015). The leader-in-social-network schema: Perceptions of network structure affect gendered attributions of charisma. *Organization Science, 26*(4), 1210-1225.

Budesheim, T. L., & DePaola, S. J. (1994). Beauty or the beast? The effects of appearance, personality, and issue information on evaluations of political candidates. *Personality and Social Psychology Bulletin, 20*(4), 339-348.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science, 6*(1), 3-5.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35*(1), 38.

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues, 57*(4), 725-741.

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin, 41*(3), 405-419.

Carpinella, C. M., Hehman, E., Freeman, J. B., & Johnson, K. L. (2016). The gendered face of partisan politics: consequences of facial sex typicality for vote choice. Political Communication, 33(1), 21-38.

Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology, 113*(6), 1479-1526.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford law review,* 1241-1299.

Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology, 72*(1), 97.

Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep. net/stat/mlelr. pdf.*

Derks, B., Ellemers, N., Van Laar, C., & De Groot, K. (2011). Do sexist organizational cultures create the Queen Bee?. *British Journal of Social Psychology*, *50*(3), 519-535.

Duehr, E. E., & Bono, J. E. (2006). Men, women, and managers: are stereotypes finally changing?. *Personnel Psychology*, *59*(4), 815-846.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Harvard Business Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The leadership quarterly*, *14*(6), 807-834.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: race, crime, and visual processing. *Journal of personality and social psychology*, *87*(6), 876.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383-386.

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13593-13597.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, *82*(6), 878.

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.

Gorman, E. H. (2005). Gender stereotypes, same-gender preferences, and organizational variation in the hiring of women: Evidence from law firms. *American Sociological Review*, *70*(4), 702-728.

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science*, *5*(7), 815-824.

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science*, *5*(7), 815-824.

Heilman, M. E., & Stopeck, M. H. (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. *Organizational Behavior and Human Decision Processes*, *35*(2), 202-215.

Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of applied psychology*, *74*(6), 935-942.

Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, 119-147.

Jæger, M. M. (2011). "A thing of beauty is a joy forever"? Returns to physical attractiveness over the life course. *Social Forces*, *89*(3), 983-1003.

Johnson, D. S., Cinnioglu, C., Ross, R., Filby, A., Gemelos, G., Hill, M., ... & Murray, M. J. (2010). Comprehensive analysis of karyotypic mosaicism between trophectoderm and inner cell mass. *Molecular human reproduction, 16*(12), 944-949.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes, 106*(1), 39-60.

Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical attractiveness biases in ratings of employment suitability: Tracking down the "beauty is beastly" effect. *The Journal of social psychology, 150*(3), 301-318.

King, A., & Leigh, A. (2009). Beautiful politicians. *Kyklos, 62*(4), 579-593.

Klofstad, C. A., Anderson, R. C., & Nowicki, S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. *PloS one, 10*(8), e0133779.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms.

Kuwabara, K., & Thébaud, S. (2017). When Beauty Doesn't Pay: Gender and Beauty Biases in a Peer-to-Peer Loan Market. *Social Forces, 95*(4), 1371-1398.

Livingston, R. W., & Pearce, N. A. (2009). The teddy-bear effect: Does having a baby face benefit black chief executive officers?. *Psychological Science, 20*(10), 1229-1236.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology, 31*(1), 41-58.

McLean, W. P., & Unter, K. A. (2010). SEX ROLE STEREOTYPING AND REQUISITE MANAGEMENT CHARACTERISTICS. *Economics, Management & Financial Markets, 5*(4).

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*(3), 455.

Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly, 25*(5), 817-834.

Parks-Stamm, E. J., Heilman, M. E., & Hearns, K. A. (2008). Motivated to penalize: Women's strategic rejection of successful women. *Personality and Social Psychology Bulletin, 34*(2), 237-247.

Perrett, D. I., Lee, K. J., Penton-Voak, I., & Rowland, D. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*(6696), 884.

Pratto, F., & Espinoza, P. (2001). Gender, ethnicity, and power. *Journal of Social Issues, 57*(4), 763-780.

Re, D. E., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2013). Facial cues to perceived height influence leadership choices in simulated war and peace contexts. *Evolutionary Psychology, 11*(1), 147470491301100109.

Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford University Press.

Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects

of organizational performance on leaders with single versus dual-subordinate identities. *Journal of Experimental Social Psychology, 48*(5), 1162-1167.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues, 57*(4), 743-762.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues, 57*(4), 743-762.

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.

Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success. *Sex Roles, 61*(9-10), 644-652.

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.

Schein, V. E., & Mueller, R. (1992). Sex role stereotyping and requisite management characteristics: A cross cultural look. *Journal of Organizational Behavior, 13*(5), 439-447.

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of applied psychology, 60*(3), 340.

Schein, V. E., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager--think male: A global phenomenon?. *Journal of organizational behavior*, 33-41.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of personality and social psychology, 45*(3), 513.

Sczesny, S., & Kühnen, U. (2004). Meta-cognition about biological sex and gender-stereotypic physical appearance: Consequences for the assessment of leadership competence. *Personality and Social Psychology Bulletin, 30*(1), 13-21.

Sczesny, S., Spreemann, S., & Stahlberg, D. (2006). Masculine= competent? Physical appearance and sex as sources of gender-stereotypic attributions. *Swiss Journal of Psychology, 65*(1), 15-23.

Silberzahn, R., & Menges, J. (2016). Reading the Face of a Leader: Women with Low Facial Masculinity Are Perceived as Competitive. *Academy of Management Discoveries, 2*(3), 272-289.

Spisak, Brian R., et al. "Facing the situation: Testing a biosocial contingency model of leadership in intergroup relations using masculine and feminine faces." *The Leadership Quarterly* 23.2 (2012): 273-280.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological science, 21*(3), 349-354.

Stoker, J. I., Garretsen, H., & Spreeuwers, L. J. (2016). The facial appearance of CEOs: Faces signal selection but not performance. *PloS one, 11*(7), e0159950.

Sczesny, S., & Stahlberg, D. (2002). The influence of gender-stereotyped perfumes on leadership attribution. *European Journal of Social Psychology, 32*(6), 815-828.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623-1626.

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.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of personality and social psychology, 47*(2), 237.

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs' facial structure predicts their firms' financial performance. *Psychological Science, 22*(12), 1478-1483.

Zebrowitz, L. A., & Montepare, J. M. (2008). First impressions from facial appearance cues. *First impressions*, 171-204.

Zell, E., & Bernstein, M. J. (2014). You may think you're right… Young adults are more liberal than they realize. *Social Psychological and Personality Science, 5*(3), 326-333.