Journal of Experimental Psychology: General

Perceived Femininity and Masculinity Contribute Independently to Facial Impressions

Neil Hester, Benedict C. Jones, and Eric Hehman

Online First Publication, October 29, 2020. http://dx.doi.org/10.1037/xge0000989

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

http://dx.doi.org/10.1037/xge0000989

Perceived Femininity and Masculinity Contribute Independently to Facial Impressions

Neil Hester
McGill University

Benedict C. Jones
University of Strathclyde

Eric Hehman
McGill University

In person perception research, femininity and masculinity are regularly conceived as 2 ends of 1 bipolar dimension. This unidimensional understanding permeates work on facial impressions, gender diagnosticity, and perceptions of LGBTQ individuals, but it is perhaps most prominent in evolutionary work suggesting that sexually dimorphic facial features (which vary along a female–male continuum) correspond directly with subjective ratings of femininity and masculinity, which in turn predict ratings of traits such as attractiveness. In this paper, we analyze 2 large face databases (the Chicago and Bogazici Face Databases) to demonstrate that femininity and masculinity are distinct dimensions in person perception. We also evaluate key theoretical assumptions surrounding femininity and masculinity in evolutionary theories of face perception. We find that sexually dimorphic features weakly correlate with each other and typically explain just 10–20% of variance in subjective ratings of femininity and masculinity. Femininity and masculinity each explain unique variance in trait ratings of attractiveness, dominance, trustworthiness, and threat. Femininity and masculinity also interact to explain unique variance in these traits, revealing *facial androgyny* as a novel phenomenon. We propose a new theoretical model explaining the link between biology, facial features, perceived femininity and masculinity, and trait ratings. Our findings broadly suggest that concepts that are "opposites" semantically cannot necessarily be assumed to be psychological opposites.

Keywords: impression formation, social cognition, androgyny, gender, evolutionary psychology

In 2017, National Geographic stated that we are in the midst of a "gender revolution." For example, the Oxford English Dictionary had just recently added "genderqueer, "gender-fluid", and the gender-neutral title "Mx" to its pages (Tan, 2016). As "woman" and "man" have lost their perceived status as natural kinds, traditional "feminine" and "masculine" appearances have also lost popularity in favor of more androgynous looks. In Western culture, the "butch chic" style that "was once a queer-owned style has shifted to the mainstream" and no longer signals sexual preference (Wilkinson, 2015). In South Korea, the "salaryman" aesthetic has given way to *Kkonminan*, a word combining "flower" and "handsome man" that describes men whose faces look "soft yet manly at the same time" (BBC, 2018). From these trends, it seems that someone can easily appear feminine and masculine at once and that our very understanding of these concepts is influenced by top-down social knowledge concerning gender. Yet, psychological research on facial impressions often assumes otherwise.

Femininity and Masculinity in Facial Impression Research

In research on person perception, perceived femininity and masculinity are regularly conceived as—and measured as—two ends of one bipolar dimension (Little, Jones, & DeBruine, 2011; Mitteroecker, Windhager, Müller, & Schaefer, 2015; Perrett et al., 1998; Rhodes, 2006; Watkins, DeBruine, Little, & Jones, 2012). This conception of femininity–masculinity[1] is rooted in a presumed theoretical link between objective sexually dimorphic facial features (which are thought to vary on a single female–male dimension) and subjective ratings of femininity and masculinity

Neil Hester, Department of Psychology, McGill University; Benedict C. Jones, School of Psychological Sciences and Health, University of Strathclyde; Eric Hehman, Department of Psychology, McGill University.

Conceived research: Neil Hester. Methodology: All authors. Data Curation: Neil Hester. Analysis: Neil Hester. Writing – Original Draft: Neil Hester. Writing – Review and Editing: All authors. This research was supported by a SSHRC Insight Development Grant (430-2016-00094) to Eric Hehman. We thank Jax Norman for her help compiling and summarizing relevant papers for the project, as well as Chevieve Heri for her help copyediting the manuscript. Our data, syntax, and supplemental materials are available at https://osf.io/yjn5w/.

Correspondence concerning this article should be addressed to Neil Hester, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada. E-mail: neilrhester@gmail.com

[1] "Femininity–masculinity" is used to describe work in which femininity and masculinity are assumed to be opposite ends of a single dimension.

(which are thought to vary on a single feminine–masculine dimension). Sexually dimorphic facial features and perceived femininity and masculinity are presumed to both strongly correspond and vary along one-dimensional bipolar axes.

This idea of "strong correspondence" between sexually dimorphic facial features and perceived femininity and masculinity features prominently as part of the immunocompetence handicap hypothesis in humans. This hypothesis argues that facial masculinity is a sexual ornament signaling reproductive advantages via resistance to disease. Testosterone suppresses the immune system, such that reproductively desirable males with stronger immune systems are theoretically able to maintain higher levels of testosterone, which in turn causes more extreme expressions of facial masculinity (Thornhill & Gangestad, 1999; see Zaidi et al., 2019 for summary and critique). Facial masculinity within males thus provides an "honest signal" of mate quality and should correspond with perceptions of attractiveness. An analogous model exists for females: Facial femininity supposedly provides an "honest signal" of estrogen levels, which is why men are attracted to more feminine faces (Law Smith et al., 2006; Penton-Voak et al., 2001).

Importantly, immunocompetence handicap models assume that objective sexually dimorphic features and subjective perceptions of femininity and masculinity tightly correspond with each other—faces are perceived as feminine or masculine because they are, objectively, feminine or masculine. This assumption is apparent in a host of earlier work on feminine and masculine faces, which measured perceived femininity and masculinity as two ends of a single dimension (e.g., Cunningham, Barbee, & Pike, 1990; Gillen, 1981; O'Toole et al., 1998; Rhodes, Chan, Zebrowitz, & Simmons, 2003) or, alternatively, measured ratings of femininity or masculinity alone, as if "low feminine" was equivalent to "high masculine" (e.g., Mitteroecker et al., 2015). Later work often focused more on "objective" femininity–masculinity by morphing faces to appear more feminine or more masculine along a single dimension (e.g., O'Connor, Fraccaro, Pisanski, Tigue, & Feinberg, 2013; Perrett et al., 1998; Rennels, Bronstad, & Langlois, 2008). Reviews of these differing approaches acknowledge the methodological differences but also assume that both approaches are capturing the same general latent factor (Rhodes, 2006).

The assumption that femininity and masculinity are two discrete ends of a single dimension extends beyond research focusing on heterosexual attractiveness and the immunocompetence handicapping hypothesis. This is illustrated by many researchers using a single item of either femininity or masculinity to capture gender-typical variation. For instance, in the impression formation literature, the femininity and masculinity of a face is often measured with a single item or measured with a bipolar scale (e.g., Hehman, Sutherland, Flake, & Slepian, 2017; Sutherland et al., 2013). Other work uses gender diagnosticity, a Bayesian probability that a person in a population is female or male (Lippa, 1995; Lippa & Arad, 1999; Lippa & Connelly, 1990; Lippa, Martin, & Friedman, 2000). This value, ranging from "definitely female" to "definitely male," is computed using various sets of gender-related indicators and predicts outcomes such as people's occupational interests (Lippa, 1995, 1998b), health (Lippa et al., 2000), and test performance (Lippa, 1998a).

Much work on femininity–masculinity is heteronormative, which is unsurprising given its roots in evolutionary theory and the relative "invisibility" of LGBTQ people for much of psychology's existence (see Clarke, Ellis, Peel, & Riggs, 2010). Research on LGBTQ people acknowledges that the "female-male" binary fails to describe many people's sexual and gender identities. However, this research still typically conceives of femininity and masculinity as two ends of a single dimension, rather than two distinct factors. Perceived femininity–masculinity serves as a signal of sexual orientation, such that targets that are perceived as less gender-normative are also more likely to be gay or lesbian (Rule, Ambady, Adams, & Macrae, 2008; Rule, Ambady, & Hallett, 2009; Rule & Alaei, 2016). This pattern has been found both using self-report measures (Rieger, Linsenmeier, Gygax, Garcia, & Bailey, 2010; Valentova, Rieger, Havlicek, Linsenmeier, & Bailey, 2011) and facial manipulations (Freeman, Johnson, Ambady, & Rule, 2010). Research on femininity–masculinity and transgender individuals shows that the same photograph is perceived as less feminine/more masculine (using a bipolar scale) when the person in the photograph "identifies as transgender" (Howansky, Albuja, & Cole, 2020). Furthermore, people evaluate transgender people more negatively when they possess physically androgynous (vs. sex-typical) bodies (Stern & Rule, 2018). Finally, researchers have also examined how self-reported femininity–masculinity relates to sexual arousal for perceivers of varied gender and sexual orientation (Rieger, Savin-Williams, Chivers, & Bailey, 2016).

Femininity and Masculinity as Distinct Concepts

There is a large body of research that conceives femininity and masculinity as two ends of one dimension. Historically, this makes sense: Traditionally humans have associated various traits, roles, and behaviors with either women or men. Despite this dichotomization, the idea of androgyny—possessing both feminine and masculine characteristics—has also existed for centuries, suggesting that people may view femininity and masculinity as distinct (if related) concepts rather than two extremes of a single dimension. Although early psychological research clustered "feminine" and "masculine" traits on opposite ends of a single dimension (Terman & Miles, 1936), two-factor models eventually replaced them, conceiving femininity and masculinity as orthogonal (see Lippa, 2001). One recent paper argues that independent dimensions of femininity and masculinity are the true basis for the fundamental "Big Two" underlying social cognition (Martin & Slepian, 2020). A two-factor understanding of femininity and masculinity allows for ideas such as psychological androgyny (Bem, 1974; for a recent example, see Juster et al., 2016), feminine and masculine cultures in STEM (Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Cheryan, Ziegler, Montoya, & Jiang, 2017), and gendered brand personality (Grohmann, 2009). In other areas of study such as feminist philosophy and sociology, femininity and masculinity are not only perceived as distinct but also multidimensional in themselves (people perform or demonstrate various masculinities and femininities; e.g., Budgeon, 2014; Paechter, 2006; Reay, 2001). Yet, this rich understanding of femininity and masculinity in other literatures is rarely reflected in research on person perception.

What does this mean for the perception of individuals? Although female and male faces certainly differ, perceivers' complex concepts of femininity and masculinity might exert top-down influence on their subjective face ratings (Freeman & Ambady, 2011). In other words, people probably do not simply think "how biologically female/male does this face look?" when they rate femininity and masculinity.

Instead, they likely account for factors such as resting facial expression (Hester, 2019; Zebrowitz, 2017), facial width-to-height ratio (Geniole, Denson, Dixson, Carré, & McCormick, 2015), overall attractiveness (Scott, Pound, Stephen, Clark, & Penton-Voak, 2010), and hairstyle (Freeman, Ambady, Rule, & Johnson, 2008). Furthermore, their interpretation of these factors likely varies as a function of their own endorsement of gender stereotypes or sex roles (Wood & Eagly, 2012; see Hehman, Stolier, Freeman, Flake, & Xie, 2019; Xie, Flake, & Hehman, 2019).

In the present research, we tested whether perceived femininity and masculinity in faces are independent dimensions that both contribute unique and meaningful variance to facial impressions. Though impressions of femininity and masculinity are studied and examined across a wide array of research areas, the original basis for the one-dimension conception of femininity–masculinity is rooted in biology and evolutionary theory, which makes statements not only about the dimensionality of femininity and masculinity but also their relation to facial features and perceived traits. For this reason, we considered femininity–masculinity in the larger context of the immunocompetence handicap hypothesis. We identify key theoretical assumptions about sexually dimorphic facial features, perceived femininity and masculinity, and their relation. Then, we discuss our empirical tests of these assumptions.

Evolutionary Theories of Sexual Selection: Key Theoretical Assumptions

A careful examination of evolutionary theories of sexual selection reveals four key assumptions about objective sexually dimorphic facial features, subjective perceptions of femininity and masculinity, and their relation. See Figure 1 for a theoretical model of the immunocompetence handicap hypothesis with the tested assumptions described in red. In the sections below, we review each assumption, consider the evidence supporting such assumptions, and describe our approach to empirically test them.

Assumption I: Dimorphic Facial Features Reflect a Single Latent Factor

Sexually dimorphic facial features are thought to serve as honest signals of hormonal differences (Little et al., 2010; Waynforth, Delwadia, & Camm, 2005). Theoretically, testosterone is an immunosuppressant that causes masculinized facial features, which signal to potential mates that the target has high fitness due to his ability to survive despite high levels of testosterone (Rhodes et al., 2003; Thornhill & Gangestad, 1999).

However, studies linking hormonal differences to facial features yield mixed results, sometimes due to small samples and/or relatively small effect sizes (Fink et al., 2005; Neave, Laing, Fink, & Manning, 2003; Penton-Voak & Chen, 2004), calling into question the idea that hormones are the "common cause" of sexually dimorphic facial features. Of course, there are challenges with linking hormones to facial development. For example, prenatal hormone levels may be more strongly linked to masculine facial features than adult hormone levels, explaining researchers' mixed results examining adult testosterone levels (e.g., Whitehouse et al., 2015).

If certain hormones are the "root cause" of sexually dimorphic features, then these sexually dimorphic facial features should correlate with each other strongly enough to reflect a single underlying factor (Assumption I). We tested Assumption I by examining the correspondence between numerous facial features identified by previous research as differing between women and men (Burriss, Welling, & Puts, 2011; Ma, Correll, & Wittenbrink, 2015; Mitteroecker et al., 2015; Penton-Voak et al., 2001). We also formally tested the fit of a model in which these sexually dimorphic features load onto a single latent factor. Finally, we used exploratory techniques to explore the actual factor structure of these facial features.

Assumption II: Perceived Femininity–Masculinity is Mostly a Reflection of Sexually Dimorphic Facial Features

The immunocompetence handicap hypothesis also suggests that sexually dimorphic facial features and perceived femininity–

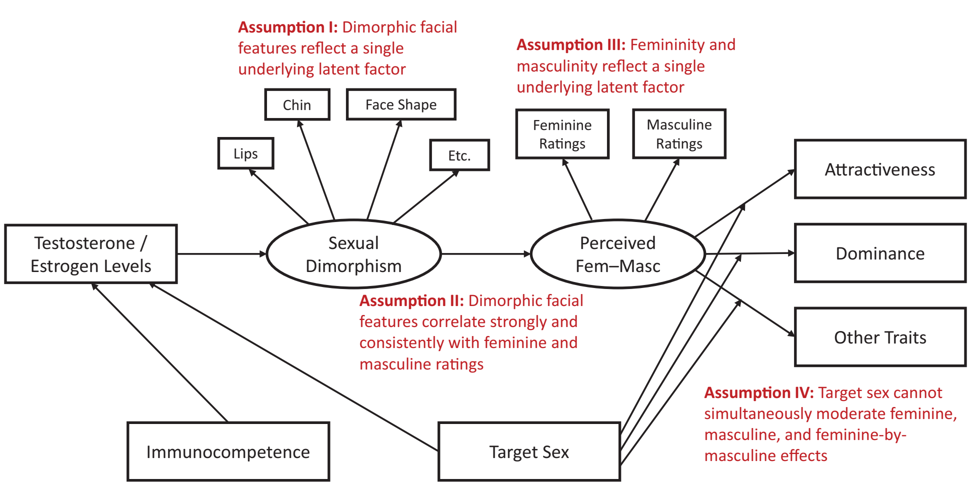

Figure 1. A structural summary of immunocompetence handicap hypothesis explaining the structure of human faces. The theory makes key assumptions that are evaluated in the present article, the most important of which is the definition of femininity and masculinity as two ends of a single dimension. See the online article for the color version of this figure.

masculinity should strongly correspond with each other (e.g., Thornhill & Gangestad, 1999; Waynforth et al., 2005). Although sexually dimorphic features are derived from differences between female and male faces (Perrett et al., 1998; Rhodes, 2006), researchers often assume that within-gender variation in these features will correspond similarly to perceived femininity–masculinity regardless of whether the target is female or male. This assumption often goes untested, with many manipulations of sexually dimorphic features lacking a manipulation check in the form of subjective feminine and masculine ratings (e.g., Burriss, Welling, et al., 2011; Fraccaro et al., 2010; Marcinkowska, Jasienska, & Prokop, 2018).

However, femininity and masculinity are rich social concepts that take on different meanings depending on the gender of the target (Martin & Slepian, 2020). Thus, the association between sexually dimorphic features and subjective ratings of femininity and masculinity may vary as a function of target gender. For example, prominent cheekbones might be more associated with masculinity for male targets (compared to female targets) and more associated with femininity for female targets (compared to male targets). In other words, the relationship between morphological features and perceptions of femininity or masculinity may be moderated by perceived target sex.

Furthermore, sexually dimorphic features may simply explain less within-gender variance in subjective evaluations of femininity and masculinity, compared to across-gender variance. Recent studies have shown a large amount of unexplained variance in perceived masculinity within male faces. Many standard combinations of sexually dimorphic facial features used to capture "maleness" only appear to explain 10–15% of the variance in ratings of masculinity within males (Sanchez-Pages, Rodriguez-Ruiz, & Turiegano, 2014), and more sophisticated morphometric methods (i.e. methods that measure specific features in the face and their relation to other features) only explain 25% of the variance in ratings of masculinity within males (Mitteroecker et al., 2015). Thus, it might be the case that most within-gender variance in ratings of femininity and masculinity is explained by factors other than sexually dimorphic features.

If perceived femininity–masculinity is simply a reflection of sexually dimorphic features, then sexually dimorphic features should predict perceived femininity and masculinity equally for male and female targets (Assumption IIa). Furthermore, sexually dimorphic features should explain equally large amounts of variance in perceived femininity and masculinity both across-gender and within-gender (Assumption IIb). To test Assumption IIa, we examined whether the relation between sexually dimorphic features and ratings of femininity and masculinity was moderated by target gender. To test Assumption IIb, we examined whether sexually dimorphic features explain less variance within-gender than they do across-gender.

Assumption III: Femininity and Masculinity are Two Ends of a Single Dimension

The prior two assumptions lead to the key third assumption about the basic structure of feminine and masculine ratings: Subjective ratings of femininity and masculinity are assumed to represent two ends of a single bipolar dimension. However, as discussed earlier, femininity and masculinity may actually constitute distinct (albeit related) dimensions.

If femininity and masculinity do constitute a single dimension, then ratings of facial femininity and masculinity should be equally and strongly negatively correlated both across-gender and within-gender (Assumption IIIa). That they are measured in different contexts (e.g., among women only, among men only) should not change the relationship between femininity and masculinity if they truly reflect one dimension.

Additionally, if femininity and masculinity reflect the same dimension, then femininity should not explain variance in outcomes above and beyond masculinity (Assumption IIIb). As an analogy, it would not make sense to predict that a model with both "tall" and "short" as predictors would explain more variance in "being a good basketball player" than a model with just "tall" in it. Finally, if femininity and masculinity reflect the same dimension, then the interaction between femininity and masculinity should not explain significant variance in outcomes (Assumption IIIc). Put differently, two items that capture the same latent factor should interact with each other to explain variance (a latent factor cannot interact with itself). To continue the earlier example, it would be strange to find that the interaction between "tall" and "short" explains even more variance in this outcome.

We tested Assumption IIIa by correlating feminine and masculine ratings across male and female faces, within male faces, and within female faces, with the prediction that within-gender correlations would be significantly weaker than across-gender correlations. To test Assumption IIIb, we examined whether models that include both feminine and masculine ratings as predictors explain significantly more variance in four outcomes—attractiveness, dominance, trustworthiness, and threat—than models including only masculine ratings. Finally, to test Assumption IIIc, we examined whether the interaction between femininity and masculinity explains variance in these outcomes above and beyond the femininity + masculinity model. If the Femininity × Masculinity interaction consistently explains unique variance, then this would provide strong evidence that femininity and masculinity do not reflect the same dimension.

Assumption IV: Target Sex Cannot Moderate Separate Effects of Femininity and Masculinity on Trait Judgments

We examined an additional assumption that is a direct consequence of conceptualizing femininity and masculinity as two ends of one dimension: that target sex cannot simultaneously moderate effects of femininity, masculinity, and Femininity × Masculinity on traits such as attractiveness (Assumption IV). Should target sex moderate how these variables predict traits, this would reveal that the relationship is not equivalent across target sex, suggesting that people's concepts of femininity and masculinity vary systematically depending on target gender. We tested this assumption by including target sex as a moderator in the models predicting traits by perceived femininity and masculinity.

Importantly, the strength of these assumptions is not clearly defined in the field, varying considerably depending on the specific work. Some of these assumptions, as stated in the literature, arguably lack explicit standards by which they can be falsified. We have done our best to create clear but fair "objective criteria" for these assumptions to allow for falsification, which is key to any scientific theory.

Method

Before describing our tests for each theoretical assumption, we describe our two sources of data. These two distinct data sets allow us to test and replicate our conclusions across different target stimuli and perceiver cultures.

Chicago Face Database Codebook

The Chicago Face Database (CFD) includes coder ratings for 597 neutrally posed faces ($M_{perceivedage}$ = 28.86 years, $SD_{\text{perceivedage}}$ = 6.30, $Range_{perceivedage}$ = 17 to 56; 51% female; 18% Asian, 33% Black, 18% Hispanic, 31% White). Coders were 64% female, 40% non-White, varied considerably in age (M = 26.8, SD = 10.5), and were mostly American. Interrater reliability of relevant ratings was high ($\alpha s > .99$).[2]

Though the CFD's primary purpose is to inform stimulus selection, it is also a rich source of facial impressions data. Subjective ratings include perceived femininity, masculinity, attractiveness, threat, trustworthiness, and dominance. Included objective measurements are various facial features such as nose width, face width, and lip thickness (measured in pixels). Coders rated target faces using seven-point scales with the following prompt: *"Now, consider the person pictured above and rate him/her with respect to other people of the same race and gender.*—**[VARIABLE]**—(1 = *Not at all*; 7 = *Extremely*)." Coders saw 10 to 15 randomly selected faces and rated each face on all traits at once, including femininity and masculinity (Ma et al., 2015).

There is no clear consensus regarding which specific facial features are sexually dimorphic. For this reason, we used multiple sources to compile a list of features that have been previously identified at least once by researchers in the field as being sexually dimorphic (for related perspectives, see Holzleitner et al., 2019; Said & Todorov, 2011). First, Burriss, Roberts, Welling, Puts, and Little (2011) provide a detailed account of sexually dimorphic features across two studies and systematically tested these features based on evidence from prior work. Second, Mitteroecker and colleagues (2015) use a morphometric approach to identify a few other traits that differ across gender. Third, Ma and colleagues (2015) conduct principal components analysis on the objective facial measurements included in the CFD and interpret one component as a "Gender" component that explains 19.9% of the variance and correlated $r = .56$ with masculinity and $r = .54$ with femininity (across-gender correlation). Using these papers as guides, we selected 13 facial features on which to focus for analysis (see Table 1).

Bogazici Face Database Codebook

The Bogazici Face Database (BFD) includes coder ratings for 264 neutrally posed faces ($M_{perceivedage}$ = 21.65, $SD_{\text{perceivedage}}$ = 1.89, $Range_{perceivedage}$ = 19 to 32; 56% female; 99% Turkish nationals). Coders were 64% female, 40% non-White, varied considerably in age (M = 26.8, SD = 10.5), and were mostly from Turkey. Interrater reliability of relevant ratings was high (ICCs > .90).

Like the CFD, the BFD is primarily used for stimulus selection but is also a source of data for facial impressions. Rated traits included perceived femininity, masculinity, attractiveness, trustworthiness, and dominance. A morphometric variable representing the objective maleness of each target's face was additionally calculated (Mitteroecker et al., 2015). Importantly, this morphometric approach was developed as a critique of existing methods of manipulating femininity–masculinity in faces. Coders rated 16 randomly selected targets (eight female and eight male) using a seven-point Likert-type scale ranging from 1 (*not at all*) to 7 (*very much*). Unlike in the CFD, these ratings were blocked by both target gender and trait. This means that coders rated faces for femininity and masculinity separately (Saribay et al., 2018).

We note that neither of these face databases are representative of the world population, nor were they rated by representative samples. However, this is unnecessary for our goal of testing assumptions. When researchers argue that sexual dimorphism played an important role in our evolutionary history, they necessarily argue that this variation would be universal. Thus, disproving these "universal" assumptions in even one cultural sample is sufficient for evidence of falsification. However, it is still valuable to test these assumptions in two independent samples, increasing confidence that any counter evidence is not due to idiosyncrasies in one sample. We describe the similarities and differences between these samples below.

Analytic Approach

We tested our hypotheses using both the CFD and the BFD. Doing this allowed us to replicate our analyses and generalize our results across two stimulus sets that vary both culturally and methodologically. Combining analyses across these two data sets would be problematic. Culturally, the CFD is multiracial and includes American targets, whereas the BFD is mostly monoracial and includes Turkish targets. Methodologically, the CFD includes ratings of femininity and masculinity made simultaneously, whereas the BFD includes femininity and masculinity ratings from separate blocks. Finding the same patterns of results in both data sets suggests that methodological or cultural variation across the two sets are not responsible for our conclusions. In particular, replicating findings in the CFD using the BFD alleviated concerns about the instructions in the CFD, which ask participants to rate faces relative to other faces of the same race and gender.

The CFD contains ratings of 597 faces, and the BFD contains ratings of 264 faces, providing ample power for testing our hypotheses, even for analyses within females and within males. In the CFD, an average of 44 coders rated each face, and in the BFD, an average of 66 coders rated each face. Previous research has revealed that this number of ratings results in stable aggregate trait estimates (Hehman, Xie, Ofosu, & Nespoli, 2018; Jones et al., 2020). By using these face ratings as the unit of analysis, we mitigated concerns about stimulus sampling and generalizability (Judd, Westfall, & Kenny, 2012; Wells & Windschitl, 1999).

Data and Code

All data and code used in our analyses are available at https://osf.io/yjn5w/.

[2] These demographics are for Version 1 of the CFD. Version 2 demographics are not available. Also, the authors of the CFD warn that the reliability ratings are somewhat inflated due to sample size.

Table 1
List of Sexually Dimorphic Facial Features Examined

Facial feature	Description	Female or male?	Relevant citations
Eye width	Distance between corners of eyes	Female	(Burriss, Roberts, et al., 2011; Penton-Voak et al., 2001)
Eye height	Distance between lower and upper eyelids	Female	(Burriss, Little, & Nelson, 2007; Burriss, Roberts, et al., 2011)
Eye shape	Eye height ÷ Eye width	Female	(Ma, Correll, & Wittenbrink, 2015)
Eye size	Eye height ÷ Face length	Female	(Ma et al., 2015)
Face width/lower face height	Width of face divided by length of face from chin to eyes	Female	(Burriss, Roberts, et al., 2011; Penton-Voak et al., 2001)
Lower face height/face height	Length of face from chin to eyes divided by length of face from chin to hairline	Female	(Burriss, Roberts, et al., 2011; Penton-Voak et al., 2001)
Lip thickness	Distance between top and bottom of lips at thickest point	Female	(Fink et al., 2005; Mitteroecker, Windhager, Müller, & Schaefer, 2015)
Cheekbone prominence #1 / Heartshapeness	Face width at most prominent part of cheek ÷ face width at mouth	Female	(Burriss, Roberts, et al., 2011; Ma et al., 2015; Penton-Voak et al., 2001)
Cheekbone prominence #2	(Face width at most prominent part of cheek − Face width at mouth) ÷ Face length	Female	(Ma et al., 2015)
Midbrow height	Distance between middle of eyebrow and hairline	Male	(Mitteroecker et al., 2015)
Nose width	Distance between outside edges of nose at the widest point	Male	(Burriss et al., 2007; Burriss, Roberts, et al., 2011; Mitteroecker et al., 2015)
Face length	Distance between the chin and the hairline	Male	(Ma et al., 2015; Re et al., 2013)
Chin length	Distance from bottom edge of lips to chin	Male	(Ma et al., 2015)

Note. The “Female or Male?” column denotes for which gender the metric is larger.

Assumption I: Sexually Dimorphic Facial Features Reflect a Single Dimension

If dimorphic facial features are caused by hormones and serve as “honest signals” of the presence of these hormones, they ought to strongly correlate with each other (see Figure 2, left for the theory-imposed predicted correlation matrix). This would be consistent with the theoretical assumption that there is an underlying factor (i.e. hormones) causing sexually dimorphic facial features. We focused on the CFD for these analyses, as the BFD did not include objective facial measurements.

To test this assumption, we estimated a correlation matrix between all the 13 identified facial features. Results reveal that the observed relationships between these measures are weaker and more inconsistent (Figure 2, right) than their theorized relationships (Figure 2, left). A few variables correlate more strongly due to interdependency, where one variable is the function of another variable (e.g., eye shape is eye height divided by eye width). Beyond these, few strong correlations exist. Feminine features (lower-left area) should negatively correlate with masculine features.[3]

We additionally conducted a confirmatory factor analysis to examine the fit of the theorized one-factor model of sexually dimorphic features. We allowed correlated residuals between: eye width, eye height, eye shape, and eye size; lip thickness and cheekbone prominence; and face length and face width/lower face height. Correlated residuals were necessary given the interrelatedness of these features in the face (e.g., eye shape is eye height divided by eye width). The fit of this theory-derived one-factor model was quite poor, $\chi^2(N = 597, 57) = 1619.37$, $p < .001$, $CFI = .84$, $RMSEA = .214$, 90% CI [.205, .223], $SRMR = .14$. Given the poor fit, we turned to an exploratory principle components analysis. While this approach does not assume underlying latent factors and instead creates linear combinations of variables to maximize variance explained (Widaman, 2018), it can be used as a method to determine if there are distinct clusters of relationships in the observed data.

If sexually dimorphic features reflect a single latent factor, then the majority of variance in a principal components analysis should be accounted for by the first component. Instead, our principal components analysis with oblimin rotation yielded a first component that only accounted for 24.3% percent of the variance (Eigenvalue = 3.16). The second component accounted for 24.0% of the variance (Eigenvalue = 3.12), almost the same amount of variance as the first component. The third component accounted for 15.3% of the variance (Eigenvalue = 1.98), the fourth component accounted for 11.8% of the variance (Eigenvalue = 1.54), and the fifth component accounted for 8.7% of the variance (Eigenvalue = 1.13). These results are inconsistent with the idea that sexually dimorphic features constitute a single latent factor.

Discussion

Analyses of sexually dimorphic facial features do not support the possibility of a single latent factor. Thus, it is unlikely that sexually dimorphic features emerge from a root cause such as testosterone or estrogen levels. Instead, it seems likely that myriad

[3] We also conducted representational similarity analysis for these correlations to compare the structure of the matrices for male targets and female targets. Matrices were similar between target sex. Code to run this analysis is provided in the syntax file.

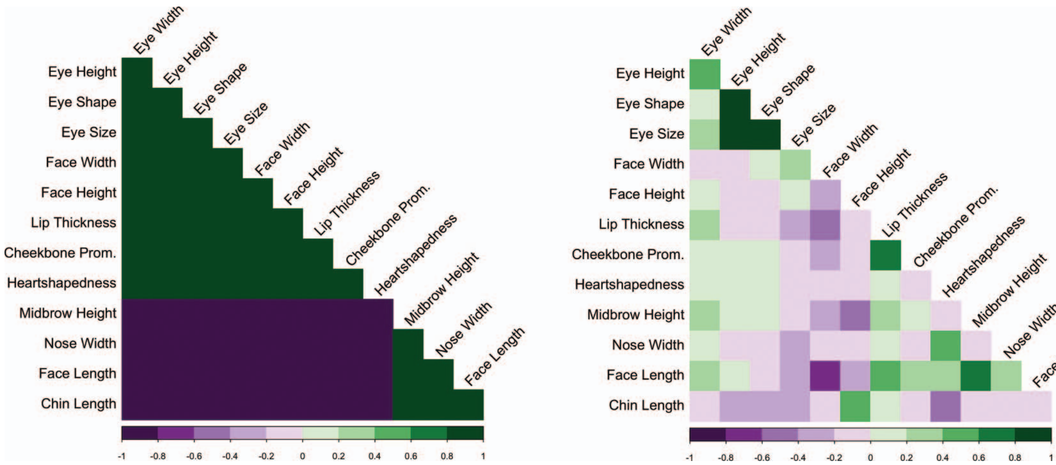

Figure 2. Correlation matrices for theoretical relationships between sexually dimorphic facial features (left) and observed relationships between sexually dimorphic facial features (right). See the online article for the color version of this figure.

biological factors (hormones, genes, nutrition, body size, etc.) determine these features. However, despite their weak correspondence with each other, it remains possible that sexually dimorphic features explain the majority of variance in ratings of femininity and masculinity both across-gender and within-gender.

Assumption IIa: Dimorphic Facial Features Predict Feminine and Masculine Ratings Consistently Across Gender

If the perceived femininity–masculinity of faces is primarily a reflection of sexually dimorphic facial features, then dimorphic features should consistently predict femininity and masculinity ratings across genders—that is, the effect of a given dimorphic feature on femininity or masculinity ratings should not be moderated by target gender, as this would suggest that people are using social knowledge or beliefs about gender to interpret these facial features.

To test this assumption, we considered whether target gender moderates the relation between sexually dimorphic features and femininity and masculinity ratings. First, we standardized all sexually dimorphic facial features to allow for comparison across features. Then, we regressed femininity and masculinity on 10 of the sexually dimorphic facial features[4] and their interactions with target gender. If perceived femininity and masculinity are mostly stable reflections of variance in sexually dimorphic facial features—without input from social knowledge about gender—then with an $\alpha = .05$ we would expect .5 of 10 tests of the interaction between dimorphic facial features and target gender to be statistically significant due to sampling variability (i.e. Type I error). In the model predicting masculinity, six of the 10 feature-by-gender interactions are significant. In the model predicting femininity, four of the 10 feature-by-gender interactions are significant. These patterns suggest that the link between sexually dimorphic facial features and perceived femininity and masculinity is not consistent across gender, providing evidence against Assumption IIa. See Figure 3 for effect sizes and Cis and Supplemental Materials (spreadsheet "SexDim") for all *F* values and *p* values.

Assumption IIb: Sexually Dimorphic Facial Features Predict the Same Percentage of Variance Across- and Within-Gender

The second part of testing Assumption II—that the perceived femininity–masculinity of faces mostly reflects sexually dimorphic facial features—is examining whether these features explain similar amounts of variance across-gender and within-gender. We test Assumption IIb using both the sexually dimorphic features in the CFD and the morphometric face shapes in the BFD.

CFD Facial Features

We regressed femininity and masculinity on the 13 facial features[5] (a) with both female and male targets, (b) with only female targets, and (c) with only male targets. Then, we statistically compared the percentages of variance explained in each model using two-tailed z-score tests for population proportions comparing the adjusted R^2 values.

Femininity. Including both female and male targets, the 13 facial features accounted for 39% of the variance in femininity. Including only female targets, these facial features only accounted for 19% of the variance in femininity, which was again a lower percentage than the across-gender model, $z = 6.09$, $p < .001$. Including only male targets, the same facial features only accounted for 25% of the variance in femininity, a lower percentage than the across-gender model, $z = 4.12$, $p < .001$.

Masculinity. Including both female and male targets, the 13 facial features accounted for 43% of the variance in masculinity. Including only female targets, these facial features only accounted for 31% of the variance in masculinity, which was again a lower

[4] The model with all thirteen facial features included as predictors showed severe collinearity issues with VIFs upwards of 800, likely due to shared values in these calculations. Removing eye shape, eye size, and face length from the models resulted in models where all VIFs < 10.

[5] Here, we include all thirteen facial features because high VIFs make individual coefficients uninterpretable but pose no issues for interpretation of the R^2 value (Aiken & West, 1991).

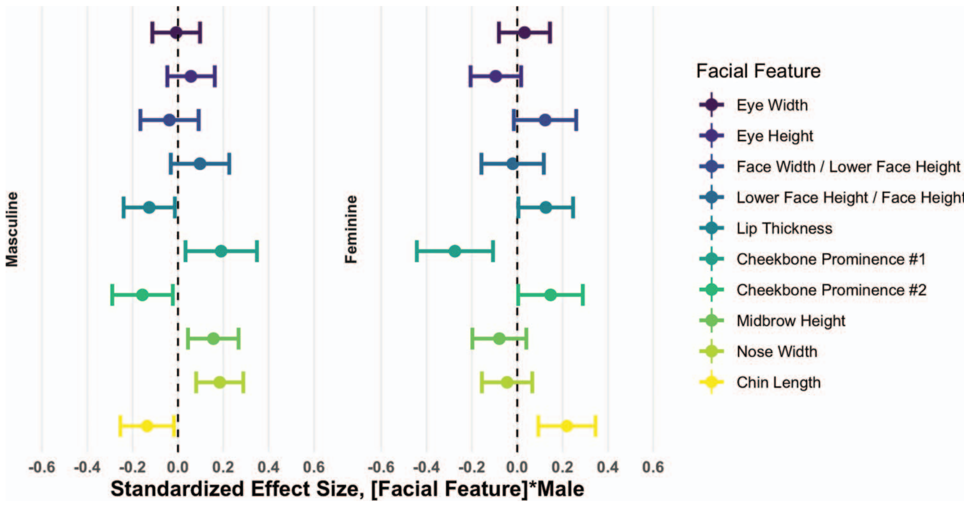

Figure 3. Sexually dimorphic facial features significantly interact with target gender to predict masculinity (six of 10 facial features) and femininity (four of 10 facial features), contrary to the theoretical assumption that ratings of femininity and masculinity reflect sexually dimorphic facial features consistently regardless of target gender. See the online article for the color version of this figure.

percentage than the across-gender model, $z = 3.50$, $p < .001$. Including only male targets, the same facial features only accounted for 27% of the variance in masculinity, a lower percentage than the across-gender model, $z = 4.61$, $p < .001$.

BFD Morphometric Male Face Shape

It is possible that a more advanced technique for capturing sexually dimorphic facial features might do a better job predicting within-gender femininity and masculinity. The BFD includes a general "male face shape" variable derived using a morphometric approach (Mitteroecker et al., 2015) that was developed to address issues with existing manipulations and measurements of sexual dimorphism in faces. This variable, according to evolutionary theories of sexual selection, ought to explain substantial amounts of variance in femininity and masculinity ratings both across-gender and within-gender. One advantage of this morphometric variable is that it accounts for complex patterns of individual facial features that might holistically reflect a "male face shape" (which cannot be adequately captured using measurements of multiple discrete facial features). Due to its holistic nature, the morphometric variable also addresses concerns in the CFD analyses that the omission of specific dimorphic facial features might be responsible for the difference between across-gender and within-gender variance explained and/or the lower-than-expected variance explained within-gender.

Femininity. Including both male and female faces, the correlation between male face shape and femininity is $r = -.72$, 95% CI $[-.78, -.67]$, such that 53.0% of the variance in femininity is explained by male face shape. However, within male faces this correlation is only $r = -.34$, 95% CI $[-.50, -.17]$, such that only 11.6% of the variance in femininity is explained by male face shape, significantly less than in the overall correlation, $z = 5.02$, $p < .001$. This correlation is also weaker within female faces, $r = -.20$, 95% CI $[-.35, -.04]$, such that only 4.0% of the variance in femininity explained by male face shape—though this does not differ significantly from the within-male outcomes, $z = 1.16$, $p = .250$. See Figure 4, right.

Masculinity. The relation between masculinity and male face shape shows similar patterns. Including both male and female faces, the correlation between male face shape and masculinity is $r = .66$, 95% CI [.59, .73], such that 43.5% of the variance in masculinity is explained by male face shape. Including only male faces, this correlation is $r = .44$, CI [.29, .58], such that only 19.4% of the variance in masculinity is explained by male face shape, significantly less than in the overall correlation, $z = 2.81$, $p = .005$. This correlation is even weaker within female faces, $r = .21$, 95% CI [.05, .36], such that only 4.4% of the variance in masculinity is explained by male face shape—even less than for male faces, $z = 2.08$, $p = .038$. See Figure 4, left.

Discussion

Across-gender, both the facial features from the CFD and the male face shape variable from the BFD explained considerable variance in femininity and masculinity ratings. Within-gender, however, these predictors explained significantly less variance. And, the effect of sexually dimorphic facial features on femininity and masculinity ratings often varied systematically by target gender, suggesting that within-gender variance in femininity and masculinity ratings are meaningfully influenced by factors other than sexually dimorphic features. Our findings in this section call into question the meaning of operationalizing within-gender "femininity" or "masculinity" by manipulating these features. In the most extreme case—the BFD male face shape only explaining 4% of the variance in ratings within females—such a manipulation hardly relates to participants' perceptions at all.

One interesting additional finding in these analyses is that the CFD and BFD facial features sometimes explained differing amounts of variance within female faces and within male faces. For example, within females, the CFD facial features explained 31% of the variance in masculine ratings but only 19% of the

Figure 4. Correlations between the maleness of face shape with masculine ratings (left) and feminine ratings (right) are weaker within females (red [light gray] circles) and within males (blue [light gray] triangles), compared to correlations including both females and males (purple [dark gray] line). Furthermore, ratings of both femininity and masculinity vary considerably across levels of male face shape for both men and women. See the online article for the color version of this figure.

variance in feminine ratings. This brings us to the central theoretical assumption that we test in this paper: that femininity and masculinity represent two ends of a single dimension. We test this assumption in three parts.

Assumption IIIa: Femininity and Masculinity Correlate Strongly Enough to Reflect a Single Factor Both Across- And Within-Gender

Femininity and masculinity are assumed to represent opposite ends of a single theoretical dimension, similar to short and tall or positive and negative. If this is the case, separate ratings of femininity and masculinity should correlate very strongly with each other and share most of their variance. At first glance, this appears to be true: For example, the correlation between femininity and masculinity reported in the CFD article is −.97 (94% shared variance; Ma et al., 2015).[6]

However, this correlation is across-gender, and we know that being categorized as male is a strong cue for ratings of masculinity and being categorized as female is a strong cue for ratings of femininity. If feminine and masculine ratings represent opposite ends of a single dimension, then they should be just equally strongly correlated within-gender as well. We tested this possibility by estimating correlations between femininity and masculinity both within female faces and within male faces.

CFD Correlations

In the CFD, the correlation between femininity and masculinity including both female and male faces is $r = -.95$, 95% CI [−.96, −.94]. However, within female faces the correlation is $r = -.82$, 95% CI [−.85, −.77]; within male faces, the correlation is $r = -.53$, 95% CI [−.61, −.45]. The female correlation is weaker than the overall correlation, $z = -10.05$, $p < .001$, as is the male correlation, $z = -17.51$, $p = .001$ (Figure 5, left panel).

BFD Correlations

In the BFD, the correlation including both female and male faces is $r = -.89$, 95% CI [−.91, −.86]. However, within female faces the correlation is $r = -.80$, 95% CI [−.85, −.74]; within male faces, the correlation is $r = -.80$, 95% CI [−.86, −.72]. Again, the female correlation is weaker than the overall correlation, $z = -3.08$, $p = .002$, as is the male correlation, $z = -2.88$, $p = .004$ (Figure 5, right panel).

Correcting for Attenuation

Correlations between two measures are inextricably linked to the reliability of each measure, as reliability provides an upward bound on the extent to which measures can correlate (Spearman, 1904), and therefore imperfect correlations are not unambiguous evidence that two measures are distinct. Two measures could capture the same latent factor but be imperfectly correlated simply due to measurement error. To address this possibility, we implemented Spearman's correction for attenuation, which estimates what a correlation would be if no measurement error existed (Spearman, 1904). Using the interrater reliabilities reported in the CFD and BFD articles, we compared the observed correlation to what we might expect given perfect measurement. We found that, even with perfect measurement, the correlations between femininity and masculinity are weak enough that we would conclude that

[6] Note that this correlation only included the original 157 faces in the CFD, rather than the full 597 faces we use from the extended CFD.

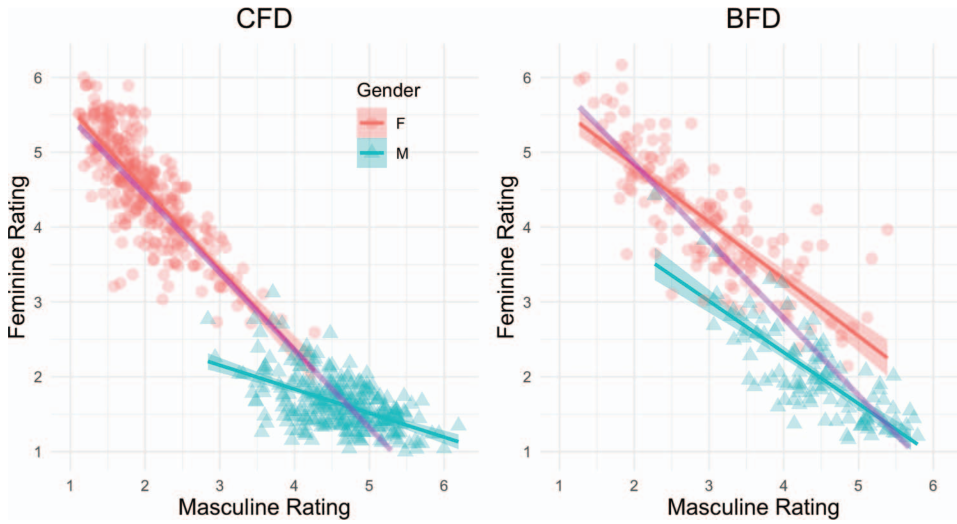

Figure 5. Correlations between ratings of femininity and masculinity are smaller when examined within females (red [light gray] circles) or within males (blue [light gray] triangles), compared to correlations including both females and males (purple [dark gray] line). As illustrated by the individual points, ratings of both femininity and masculinity vary considerably with females and males. CFD = Chicago Face Database; BFD = Bogazici Face Database. See the online article for the color version of this figure.

they are distinct factors. Within the CFD database, the femininity–masculinity relationship for female faces would change from $r = -.820$ to $r = -.821$; for male faces, from $r = -.530$ to $r = -.531$. Within the BFD database, the relation would change from $r = -.800$ to $r = -.851$ for both female and male targets. These tests suggest that our conclusions that femininity and masculinity are distinct factors cannot be explained by measurement error.

Assumptions IIIb and IIIc: Femininity and Masculinity in the Same Model Should Not Predict Unique Variance Additively or Interactively

The within-gender correlations between rated femininity and masculinity are small enough that these variables may represent distinct (if correlated) factors. However, given that we still observed fairly large correlations for some of the within-gender correlations, this is not conclusive evidence that femininity and masculinity reflect two distinct factors. For this reason, we also tested the predictive validity of femininity and masculinity as distinct factors. To do this, we estimated three hierarchical regression models for the following outcomes: attractiveness (CFD and BFD), trustworthiness (CFD and BFD), dominance (CFD and BFD), and threat (CFD). In model one, trait inferences were regressed on masculinity alone. In model two, trait inferences were regressed on both masculinity and femininity (as additive predictors; Assumption IIIb). In model three, trait inferences were regressed on masculinity, femininity, and the Masculinity × Femininity interaction (Assumption IIIc). Our goal was to examine variance explained above and beyond the predictors estimated in the previous model. To make sure that increases in variance explained are not an artifact of including more variables in the model, we use *Adjusted* R^2 as our metric and compare models for significant improvements in fit. Significant changes in *Adjusted* R^2 between models two and three suggest that masculinity and femininity are distinct and meaningful factors.

R-Squared Change Analyses

We fit three-step hierarchical regression models for each of the seven outcomes available across the CFD and BFD. For each outcome, we split the data and fit individual models for male and female targets, resulting in 14 total models. This within-gender approach provided clearer estimates of *Adjusted* R^2 and change in *Adjusted* R^2, given the clustering in masculinity and femininity ratings (see Figure 6).

With $\alpha = .05$, we would expect .7 of 14 tests of changed in R-squared to be statistically significant due to sampling variability (i.e. Type I error). Instead, we found that adding feminine ratings as a predictor alongside masculine ratings (Model 2) explained additional variance for 9 of the 14 outcomes. We also found that adding the interaction between masculine and feminine ratings (Model 3) also explained significant additional variance for 9 of the 14 outcomes, with all 7 of the male target outcomes showing additional variance explained by the Masculine × Feminine interaction. This number of significant outcomes is well above what one would expect due to sampling variability. See Supplemental Materials (spreadsheet "Rsquared") for the *F*- and *p* values for change in *Adjusted R-squared.*

Discussion

We found that femininity and masculinity were less strongly correlated within females and within males than they were across gender. However, in some cases, these correlations were still fairly large, such that the ratings may or may not have reflected two distinct factors. For this reason, we tested the unique predictive validity of femininity and masculinity and found that both ratings uniquely predicted variance in several trait inferences.

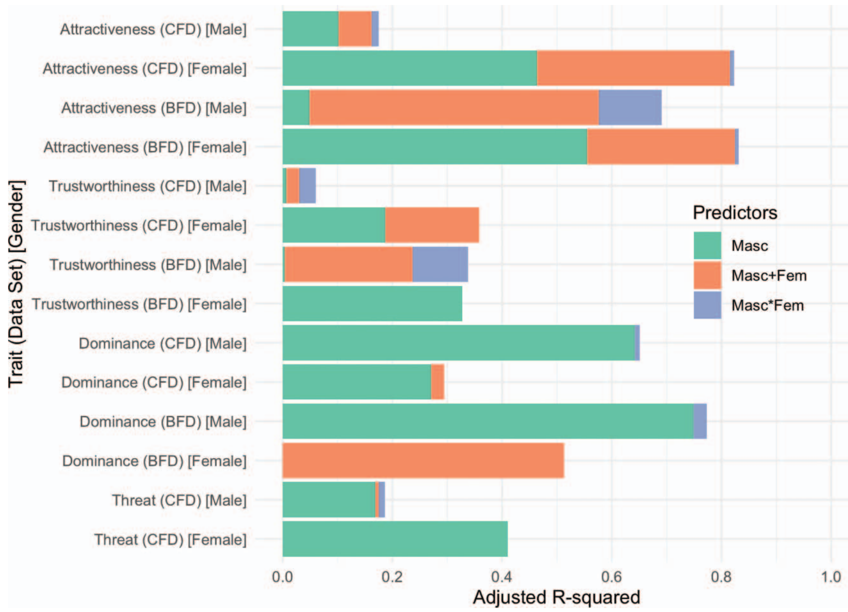

Figure 6. The R^2 values by trait, data set, and target sex. CFD = Chicago Face Database; BFD = Bogazici Face Database. See the online article for the color version of this figure.

We also found that including the Masculine × Feminine interaction in the model also explained additional variance. The prevalence of these Masculine × Feminine interactions represents especially strong evidence that the two are distinct, as two ratings that capture the same dimension should not consistently interact with each other to explain variance in outcomes. In the final section of our data analysis, we consider whether these interactive effects—along with additive effects—systematically vary by target gender.

Assumption IV: Target Gender Cannot Simultaneously Moderate Feminine, Masculine, and Feminine × Masculine Effects

Researchers have found that certain femininity–masculinity effects are sex-moderated. For example, femininity is more strongly associated with attractiveness for female than male targets (Little et al., 2011; Rhodes, 2006) and masculinity seems to be more strongly associated with dominance for men than women (Geniole et al., 2015). However, if femininity, masculinity, and the Femininity × Masculinity interaction all explain unique variance in trait inferences, then it follows that these effects might be independently moderated by target sex—a pattern that would not be predicted by prior theories in evolutionary psychology and facial impressions, as these independent moderating effects make no sense if femininity and masculinity represent two ends of one dimension.

To test for moderation of effects by target sex, we estimated models regressing our seven outcomes on femininity, masculinity, and target sex (see Figure 7). See Supplemental Materials (spreadsheets “Fem-Sex”, “Masc-Sex”, and “FemMasc-Sex”) for *t*- and *p* values.

Discussion

For different outcomes in both data sets, target sex moderated the effects of femininity, masculinity, and the Femininity × Masculinity interaction. These results bolster our primary conclusion that impressions of femininity and masculinity are distinct and meaningful sources of information. Femininity × Masculinity interactions showed a consistent pattern for male faces: Higher femininity made masculinity more strongly associated with attractiveness and trustworthiness and more weakly associated with dominance and threat. This pattern of findings suggests that male targets benefit from “facial androgyny” both additively and interactively, which differentiates them from female targets. Overall, participants’ understanding of femininity and masculinity appears to vary by target sex, suggesting that they may be incorporating stereotypical beliefs about gender into their impressions in a top-down fashion (Freeman & Ambady, 2011). See Supplementary Materials “Assumption #4 Graphs” for visualization of these results.

This result is broadly consistent with other evidence indicating that the process of forming an impression is different for female and male targets (Xie, Flake, & Hehman, 2019). Given the centrality of attractiveness ratings to evolutionary theories of sexual selection, we interpret the effects of femininity and masculinity on this outcome specifically. Overall, masculine ratings only predicted attractiveness for men (Sex × Masculinity effect). However, feminine ratings predicted attractiveness for both men and women. Furthermore, we found a clear Masculine × Feminine interaction for male targets, such that being relatively high in both masculine and feminine ratings resulted in the highest attractiveness ratings. These effects

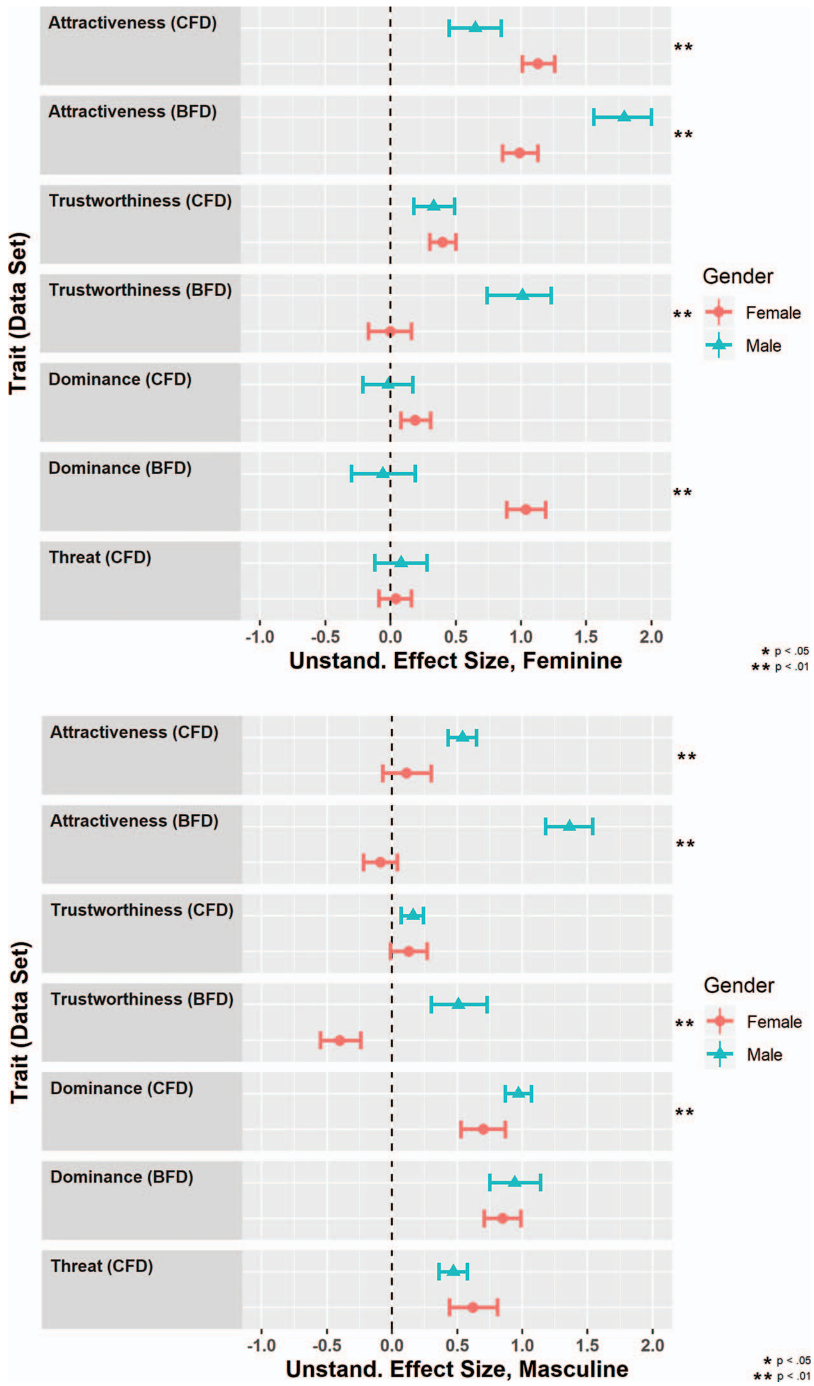

Figure 7. Unstandardized effect sizes for unique effects of femininity (top), masculinity (middle), and Femininity × Masculinity (bottom) by gender. Errors bars indicate 95% confidence intervals. CFD = Chicago Face Database; BFD = Bogazici Face Database. See the online article for the color version of this figure.

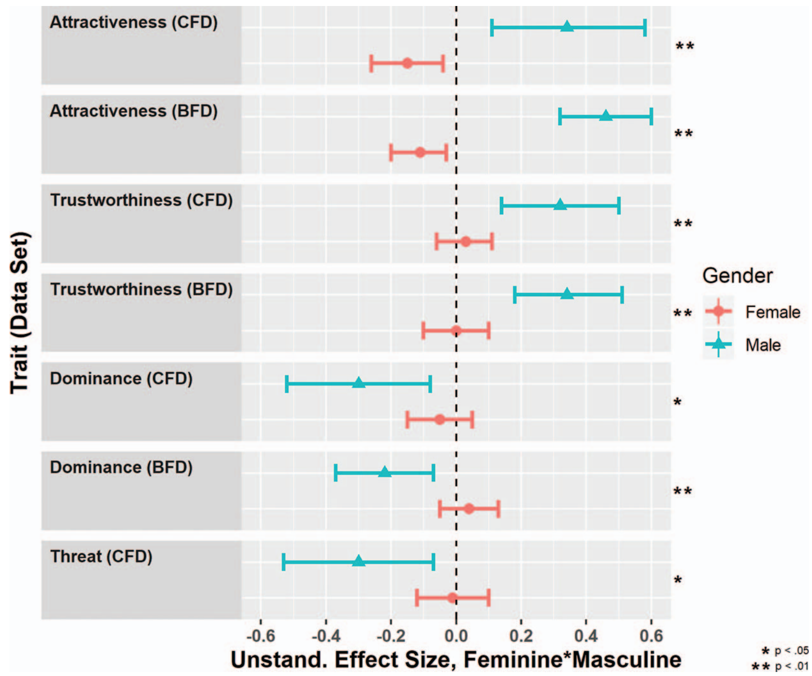

Figure 7. (*continued*)

contribute to debates over whether women find masculine- or feminine-looking men to be more attractive—the answer may simply be both.[7]

General Discussion

In contrast with decades of theory, we found that sexually dimorphic facial features only weakly correlate, suggesting that they cannot be explained by a common underlying cause (Assumption I). We also found that sexually dimorphic facial features predict perceived femininity and masculinity differently for women and men (Assumption IIa) and that these facial features predict femininity and masculinity more weakly within-gender than across-gender (Assumption IIb), suggesting that most variance in perceived femininity and masculinity is explained by other factors. We further found that perceived femininity and masculinity correlate less strongly within-gender than across-gender (Assumption IIIa) and that femininity and masculinity explain unique variance in trait ratings additively (Assumption IIIb) and interactively (Assumption IIIc), suggesting that they represent two unique sources of information rather than two ends of a single dimension. This latter set of findings challenges basic assumptions in facial impressions about the meaning of perceived femininity and masculinity in faces.

We additionally found that both femininity and masculinity uniquely predict attractiveness for men, whereas only femininity predicted attractiveness for women (Assumption IV). Finally, considering femininity and masculinity as distinct dimensions also allowed us to find potential benefits of "facial androgyny" for men: Male targets that were relatively high in both femininity and masculinity received the highest ratings in attractiveness and trustworthiness. These findings provide new perspective on debates about the role of femininity and masculinity in judgments of attractiveness. In particular, they lend credence to the lay concept of looking "androgynous" and its consequences for judgments of faces. As popular discourse rapidly acknowledges the fluidity of both gender identity and sexual orientation, psychologists should consider testing and updating their theoretical perspectives accordingly.

Establishing these dimensions as distinct also provides support for a recent theory suggesting that gendered perceptions underlie the "Big Two" structure of many models of social cognition and social impressions (Martin & Slepian, 2020). This theory is only feasible if femininity and masculinity exist as distinct dimensions. More broadly, these results highlight how concepts that are "opposites" semantically are not necessarily psychological opposites. It is often tempting to define bipolar dimensions in the interest of parsimony, and it intuitively makes sense that concepts such as "happy" and "sad" or "like" and "dislike" are polar opposites. Historically, the majority of research on emotions and attitudes has defined bipolar dimensions when examining these factors. However, research on mixed emotions shows that people can, in fact, experience "bittersweet" emotions (Larsen, Hershfield, Stastny, & Hester, 2017; Larsen, McGraw, & Cacioppo, 2001), and people's attitudes toward both objects and others can simultaneously in-

[7] Femininity predicted attractive more strongly for females (compared to males) in the CFD and more strongly for males (compared to females) in the BFD. This pattern of findings might reflect cultural differences or methodological differences. CFD males show a standard deviation (SD) of .35 for femininity ratings. BFD males show a SD of .62 for femininity ratings. On the other hand, CFD females and BFD females show similar SDs on femininity ratings, .72 and .85 respectively. The increased variance in scores might contribute to effect size differences.

clude liking and disliking (Cacioppo, Gardner, & Berntson, 1997; Cacioppo & Berntson, 1994). Here, we illustrate that variables that strongly correlate nevertheless explain unique variance in outcomes and even interact with each other to explain unique variance. When in doubt, researchers might consider the unique predictive validity of seemingly opposite variables to clarify whether these variables reflect the same dimension or different dimensions.

A Revised Model of Gendered Facial Impression

The enduring appeal of the immunocompetence handicap hypothesis comes partly from its ambitious attempt to link basic biological factors (e.g., genes, hormones) with psychological outcomes (e.g., perceived femininity, masculinity, and attractiveness) that vary by social context (e.g., mating strategy, culture; Schaefer, Mitteroecker, Fink, & Bookstein, 2009). These highly interdisciplinary models are valuable tools for integrating findings from different fields to more comprehensively understand human judgments and behavior. Based on the findings in our paper, we offer a revised model of gendered face perception to place our present findings within a theoretical framework and highlight testable hypotheses for future research (see Figure 8).

This model builds on social–cognitive models of impressions in which top-down cognitive factors constrain the manner in which bottom-up facial features inform impressions (Freeman, Stolier, & Brooks, 2020; Freeman & Ambady, 2011). It integrates this theoretical process with the model described in Figure 1, in which biology is presumed to shape the morphological features of the face. Yet here, "biological variables" is broadly defined, and facial features are not described by any latent structure. Future biological research will help clarify these links. "Facial features" are also broadly defined and—rather than being caused by target sex—instead facilitate the sex categorization of targets, which predicts feminine, masculine, and trait ratings. Furthermore, perceivers' beliefs about sex and gender exert top-down influence on numerous other pathways and even moderate the sex categorization process. As demonstrated in our studies, the model's separation of femininity and masculinity into distinct dimensions offers a more accurate picture of how these judgments predict trait inferences as well as how these effects are moderated by target sex.

We do not provide empirical evidence for the causal path from femininity and masculinity to trait judgments. We propose these causal paths by drawing on the recent perspective offered by Martin and Slepian (2020), who argue that femininity–masculinity are the innate and fundamental dimensions that underlie the "Big Two" factor structures, argued by others to be competence/warmth (Fiske, Cuddy, Glick, & Xu, 2002), agency/communion (Wiggins, 1991), agency/patiency (Schein & Gray, 2018), and valence/dominance (Oosterhof & Todorov, 2008), among others. Nevertheless, the directionality of these relations is an open question for future research. As an example, it could be the case that specific facial features cue dominance, which in turn causes judgments of high masculinity. We also do not provide empirical evidence that perceivers' endorsement of gender stereotypes would moderate the effects of perceived target sex. However, recent research supports the inclusion of these top-down attitudes in the model (e.g., Oh, Dotsch, Porter, & Todorov, 2020), and future work might also better define how gender stereotypes moderate the effect of target sex on masculine → trait and feminine → trait pathways.

In the model, we do not define a latent factor structure for specific facial features because evidence for latent factors is insufficient. However, data-driven approaches to modeling the face space using computer-generated faces provide a useful framework for understanding for understanding the effects of variation in global shape and global reflectance (Oh et al., 2020; Oosterhof & Todorov, 2008; Said & Todorov, 2011). Work using this approach has found that male faces' attractiveness is positively predicted by both feminine shape and masculine reflectance ("feminine" and "masculine" here describe dimorphic differences; Said & Todorov, 2011) and that these effects are moderated by perceivers' preferences for traits that are perceived as masculine (e.g., dominance) or

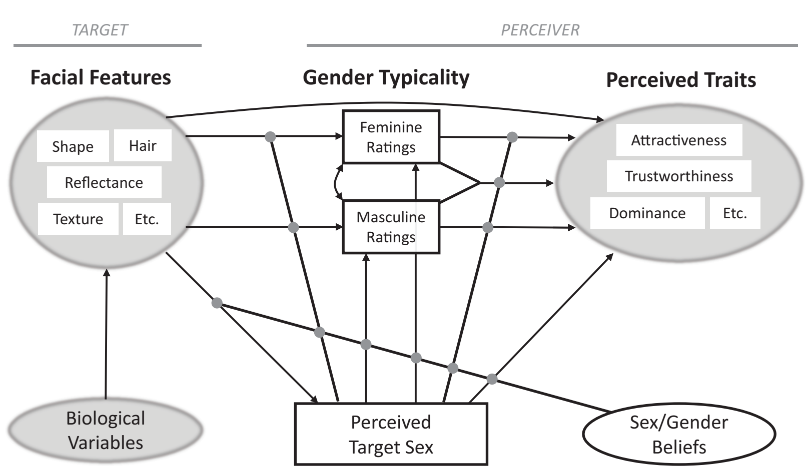

Figure 8. Proposed theoretical model linking facial features, judgments of femininity and masculinity, and perceived traits. The variables in the first column occur at the target level, whereas the variables in the second and third column occur at the perceiver level. Gray clouds represent categories; rectangles represent observed variables; and ovals represent latent variables. Arrows represent directional effects and lines with gray nodes indicate moderation.

feminine (e.g., warmth; Oh, Grant-Villegas, & Todorov, 2020). Future work might consider the extent to which these shape and reflectance dimensions account for the facial androgyny effects that we observed, in which the most positively evaluated male faces are high in both perceived femininity and perceived masculinity.

Future work might also better define the link between specific facial features and both identifications of target sex and perceptions of femininity and masculinity. The theoretical distinction between "categorization" and "evaluation" processes is an important aspect of dynamic theories of face perception (see Kawakami, Amodio, & Hugenberg, 2017). Because of this work, the present model includes both sex categorization and femininity/masculinity evaluation as distinct pathways; however, the extent to which different facial features contribute to these different processes is unclear.

Limitations

Due to the correlational nature of the analyses, the present study was unable to establish a causal link between femininity/masculinity ratings and trait judgments. Furthermore, because the data we used was formatted for use as codebooks for face stimuli, we were unable to model participant-level variance in femininity/masculinity and trait judgments. Participant-level data would allow us to potentially model effects of gender stereotype endorsement and participant gender and race, as well as simply consider the amount of variance in femininity/masculinity and trait judgments that exists at the levels of the perceiver and the target (see Xie et al., 2019). However, modeling participant-level variability is not necessary for our conclusions. We test assumptions that are ostensibly universal and should apply across perceivers. Our findings allow us to safely conclude that these "universal" assumptions are incorrect, regardless of whether they are moderated by unknown perceiver characteristics.

This is not to say that we do not value generalizing findings across cultures. To this point, we analyzed data sets from two different cultures (the United States and Turkey). We nevertheless hesitate to make universalist claims about the structure of gendered person perception. For example, femininity and masculinity may be understood differently in East or South Asian cultures (Taga, 2005), American Indian cultures (Jacobs, Thomas, & Lang, 1997), and others. Future work might test the present findings in different cultural contexts—in particular, with perceivers from different cultures—to better understand how and whether the present findings change.

Finally, the codebooks that we analyzed did not include certain traits commonly associated with femininity and masculinity, such as sexual orientation (Rule & Alaei, 2016) and competence (Oh, Buck, & Todorov, 2019). Competence, in particular, is strongly associated with attractiveness (e.g., Eagly, Ashmore, Makhijani, & Longo, 1991; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013), which is mostly explained by femininity for women but is explained by both femininity and masculinity for men. Controlling for attractiveness, masculinity linearly predicts competence in men but not women (Oh, Buck, et al., 2019). Facial androgyny may positively influence competence judgments of both men and women—for men, because androgyny predicts attractiveness, and for women, because facial androgyny may help them meet unfair social expectations to be both warm and competent at once (Cuddy, Fiske, & Glick, 2004; Hoyt, 2010).

Conclusion

"Androgyny" as a concept assumes that femininity and masculinity are two distinct—if related—dimensions. This assumption is at odds with various theoretical perspectives in person perception. Here, we find evidence for androgyny in faces: Femininity, masculinity, and their interaction uniquely contribute to impressions of faces. By grounding this effect within larger evolutionary theories of sexual selection, we also demonstrated that these perceptions of femininity and masculinity are complex and certainly unable to be reduced to sexually dimorphic features. In the past decade or so, popular conceptions of gender have become far more nuanced than before. It seems appropriate, then, that we should incorporate this same nuance in our psychological theories of person perception.

Context

The main idea occurred to the lead author from the observation that lay concepts of femininity and masculinity (which include ideas like genderfluidity and androgyny) appeared to be at odds with theoretical concepts in evolutionary psychology and person perception. Further reading on femininity and masculinity revealed a larger opportunity to evaluate various key assumptions in evolutionary theory on sexual selection and facial femininity and masculinity. This article fits into Hester's research on person perception, gender, and intersectionality; Jones's research on person perception and evolutionary theory; and Hehman's research on intergroup perception. We want to emphasize the broader implication that concepts that are "opposites" semantically cannot necessarily be assumed to be psychological opposites. Semantic opposites are present throughout psychology, but being semantic opposites is not sufficient "face validity" that these opposites comprise two ends of one dimension. We believe that researchers often underemphasize careful measurement in research (see https://psyarxiv.com/hs7wm/for a recent critique), and thinking closely about the dimensionality of related concepts is an important early step for good measurement.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

BBC. (2018, September 5). "Flowerboys" and the appeal of "soft masculinity." *BBC News*. Retrieved from https://www.bbc.com/news/world-asia-42499809

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162. http://dx.doi.org/10.1037/h0036215

Budgeon, S. (2014). The dynamics of gender hegemony: Femininities, masculinities and social change. *Sociology, 48,* 317–334. http://dx.doi.org/10.1177/0038038513490358

Burriss, R. P., Little, A. C., & Nelson, E. C. (2007). 2D:4D and sexually dimorphic facial characteristics. *Archives of Sexual Behavior, 36,* 377–384. http://dx.doi.org/10.1007/s10508-006-9136-1

Burriss, R. P., Roberts, S. C., Welling, L. L. M., Puts, D. A., & Little, A. C. (2011). Heterosexual romantic couples mate assortatively for facial symmetry, but not masculinity. *Personality and Social Psychology Bulletin, 37,* 601–613. http://dx.doi.org/10.1177/0146167211399584

Burriss, R. P., Welling, L. L. M., & Puts, D. A. (2011). Men's attractiveness predicts their preference for female facial femininity when judging for short-term, but not long-term, partners. *Personality and Individual Differences, 50,* 542–546. http://dx.doi.org/10.1016/j.paid.2010.11.022

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401–423. http://dx.doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1,* 3–25. http://dx.doi.org/10.1207/s15327957pspr0101_2

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Social Psychological and Personality Science, 2,* 656–664. http://dx.doi.org/10.1177/1948550611405218

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143,* 1–35. http://dx.doi.org/10.1037/bul0000052

Clarke, V., Ellis, S. J., Peel, E., & Riggs, D. W. (2010). *Lesbian, gay, bisexual, trans and queer psychology: An introduction*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511810121

Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues, 60,* 701–718. http://dx.doi.org/10.1111/j.0022-4537.2004.00381.x

Cunningham, M. R., Barbee, A. P., & Pike, C. L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59,* 61–72. http://dx.doi.org/10.1037/0022-3514.59.1.61

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

Fink, B., Grammer, K., Mitteroecker, P., Gunz, P., Schaefer, K., Bookstein, F. L., & Manning, J. T. (2005). Second to fourth digit ratio and face shape. *Proceedings Biological Sciences, 272,* 1995–2001. http://dx.doi.org/10.1098/rspb.2005.3179

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

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology, 8,* 447–461. http://dx.doi.org/10.1177/147470491000800311

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

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137,* 673–690. http://dx.doi.org/10.1037/a0013875

Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin, 36,* 1318–1331. http://dx.doi.org/10.1177/0146167210378755

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology, 61,* 237–287. http://dx.doi.org/10.1016/bs.aesp.2019.09.005

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS ONE, 10,* e0132726. http://dx.doi.org/10.1371/journal.pone.0132726

Gillen, B. (1981). Physical attractiveness: A determinant of two types of goodness. *Personality and Social Psychology Bulletin, 7,* 277–281. http://dx.doi.org/10.1177/014616728172015

Grohmann, B. (2009). Gender dimensions of brand personality. *Journal of Marketing Research, 46,* 105–119. Scopus. http://dx.doi.org/10.1509/jmkr.46.1.105

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass, 13,* e12431. http://dx.doi.org/10.1111/spc3.12431

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113,* 513–529. http://dx.doi.org/10.1037/pspa0000090

Hehman, E., Xie, S. Y., Ofosu, E. K., & Nespoli, G. A. (2018). Assessing the point at which averages are stable: A tool illustrated in the context of person perception. *PsyarXiv*. http://dx.doi.org/10.31234/osf.io/2n6jq

Hester, N. (2019). Perceived negative emotion in neutral faces: Gender-dependent effects on attractiveness and threat. *Emotion, 19,* 1490–1494. http://dx.doi.org/10.1037/emo0000525

Holzleitner, I. J., Lee, A. J., Hahn, A. C., Kandrik, M., Bovet, J., Renoult, J. P., . . . Jones, B. C. (2019). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance, 45,* 1589–1595. http://dx.doi.org/10.1037/xhp0000685

Howansky, K., Albuja, A., & Cole, S. (2020). Seeing gender: Perceptual representations of transgender individuals. *Social Psychological & Personality Science, 11,* 474–482. http://dx.doi.org/10.1177/1948550619875143

Hoyt, C. L. (2010). Women, men, and leadership: Exploring the gender gap at the top. *Social and Personality Psychology Compass, 4,* 484–498. http://dx.doi.org/10.1111/j.1751-9004.2010.00274.x

Jacobs, S.-E., Thomas, W., & Lang, S. (1997). *Two-spirit people: Native American gender identity, sexuality, and spirituality*. Champaign: University of Illinois Press.

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M., Antfolk, J., Arinze, N., . . . Coles, N. A. (2020). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*. Advance online publication. http://dx.doi.org/10.31234/osf.io/n26dy

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54–69. http://dx.doi.org/10.1037/a0028347

Juster, R.-P., Pruessner, J. C., Desrochers, A. B., Bourdon, O., Durand, N., Wan, N., . . . Lupien, S. J. (2016). Sex and gender roles in relation to mental health and allostatic load. *Psychosomatic Medicine, 78,* 788–804. http://dx.doi.org/10.1097/PSY.0000000000000351

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. *Advances in Experimental Social Psychology, 55,* 1–80. http://dx.doi.org/10.1016/bs.aesp.2016.10.001

Larsen, J. T., Hershfield, H. E., Stastny, B. J., & Hester, N. (2017). On the relationship between positive and negative affect: Their correlation and their co-occurrence. *Emotion, 17,* 323–336. http://dx.doi.org/10.1037/emo0000231

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81,* 684–696. http://dx.doi.org/10.1037/0022-3514.81.4.684

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., . . . Hillier, S. G. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings Biological Sciences, 273,* 135–140. http://dx.doi.org/10.1098/rspb.2005.3296

Lippa, R. (1995). Do sex differences define gender-related individual differences within the sexes? Evidence from three studies. *Personality and Social Psychology Bulletin, 21,* 349–355. http://dx.doi.org/10.1177/0146167295214005

Lippa, R. (1998a). Gender-related individual differences and National Merit test performance: Girls who are '"masculine"'and boys who are '"feminine"'tend to do better. In L. Ellis & L. Ebertz (Eds.), *Males, Females, and Behavior: Toward Biological Understanding* (pp. 177–193). Westport, CN: Praeger.

Lippa, R. (1998b). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology, 74,* 996–1009. http://dx.doi.org/10.1037/0022-3514.74.4.996

Lippa, R. (2001). On deconstructing and reconstructing masculinity–femininity. *Journal of Research in Personality, 35,* 168–207. http://dx.doi.org/10.1006/jrpe.2000.2307

Lippa, R., & Arad, S. (1999). Gender, personality, and prejudice: The display of authoritarianism and social dominance in interviews with college men and women. *Journal of Research in Personality, 33,* 463–493. http://dx.doi.org/10.1006/jrpe.1999.2266

Lippa, R., & Connelly, S. (1990). Gender diagnosticity: A new Bayesian approach to gender-related individual differences. *Journal of Personality and Social Psychology, 59,* 1051–1065. http://dx.doi.org/10.1037/0022-3514.59.5.1051

Lippa, R., Martin, L. R., & Friedman, H. S. (2000). Gender-related individual differences and mortality in the Terman longitudinal study: Is masculinity hazardous to your health? *Personality and Social Psychology Bulletin, 26,* 1560–1570. http://dx.doi.org/10.1177/01461672002612010

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

Little, A. C., Saxton, T. K., Roberts, S. C., Jones, B. C., Debruine, L. M., Vukovic, J., . . . Chenore, T. (2010). Women's preferences for masculinity in male faces are highest during reproductive age range and lower around puberty and post-menopause. *Psychoneuroendocrinology, 35,* 912–920. http://dx.doi.org/10.1016/j.psyneuen.2009.12.006

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47,* 1122–1135. http://dx.doi.org/10.3758/s13428-014-0532-5

Marcinkowska, U. M., Jasienska, G., & Prokop, P. (2018). A comparison of masculinity facial preference among naturally cycling, pregnant, lactating, and post-menopausal women. *Archives of Sexual Behavior, 47,* 1367–1374. http://dx.doi.org/10.1007/s10508-017-1093-3

Martin, A. E., & Slepian, M. L. (2020). The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. *Perspectives on Psychological Science*. Advance online publication. http://dx.doi.org/10.1177/1745691620904961

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of "masculinity" in human faces. *PLoS ONE, 10,* e0118374. http://dx.doi.org/10.1371/journal.pone.0118374

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society B: Biological Sciences, 270,* 2167–2172. http://dx.doi.org/10.1098/rspb.2003.2502

O'Connor, J. J. M., Fraccaro, P. J., Pisanski, K., Tigue, C. C., & Feinberg, D. R. (2013). Men's preferences for women's femininity in dynamic cross-modal stimuli. *PLoS ONE, 8,* e69531. http://dx.doi.org/10.1371/journal.pone.0069531

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions from faces. *Psychological Science, 30,* 65–79. http://dx.doi.org/10.31234/osf.io/xy5e6

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General, 149,* 323–342. http://dx.doi.org/10.1037/xge0000638

Oh, D., Grant-Villegas, N., & Todorov, A. (2020). The eye wants what the heart wants: Female face preferences are related to partner personality preferences. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. http://dx.doi.org/10.1037/xhp0000858

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition, 26,* 146–160. http://dx.doi.org/10.3758/BF03211378

Paechter, C. (2006). Masculine femininities/feminine masculinities: Power, identities and gender. *Gender and Education, 18,* 253–263. http://dx.doi.org/10.1080/09540250600667785

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior, 25,* 229–241. http://dx.doi.org/10.1016/j.evolhumbehav.2004.04.003

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences, 268,* 1617–1623. http://dx.doi.org/10.1098/rspb.2001.1703

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., . . . Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394,* 884–887. http://dx.doi.org/10.1038/29772

Re, D. E., Hunter, D. W., Coetzee, V., Tiddeman, B. P., Xiao, D., DeBruine, L. M., . . . Perrett, D. I. (2013). Looking like a leader-facial shape predicts perceived height and leadership ability. *PLoS ONE, 8,* e80957. http://dx.doi.org/10.1371/journal.pone.0080957

Reay, D. (2001). "Spice Girls", "Nice Girls", "Girlies", and "Tomboys": Gender discourses, girls' cultures and femininities in the primary classroom. *Gender and Education, 13,* 153–166. http://dx.doi.org/10.1080/09540250120051178

Rennels, J. L., Bronstad, P. M., & Langlois, J. H. (2008). Are attractive men's faces masculine or feminine? The importance of type of facial stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 884–893. http://dx.doi.org/10.1037/0096-1523.34.4.884

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* 199–226. http://dx.doi.org/10.1146/annurev.psych.57.102904.190208

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

Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity-femininity. *Archives of Sexual Behavior, 39,* 124–140. http://dx.doi.org/10.1007/s10508-008-9405-2

Rieger, G., Savin-Williams, R. C., Chivers, M. L., & Bailey, J. M. (2016). Sexual arousal and masculinity-femininity of women. *Journal of Per-*

sonality and Social Psychology, 111, 265–283. http://dx.doi.org/10.1037/pspp0000077

Rule, N. O., & Alaei, R. (2016). "Gaydar": The perception of sexual orientation from subtle cues. *Current Directions in Psychological Science, 25,* 444–448. http://dx.doi.org/10.1177/0963721416664403

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology, 95,* 1019–1028. http://dx.doi.org/10.1037/a0013194

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology, 45,* 1245–1251. http://dx.doi.org/10.1016/j.jesp.2009.07.010

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science, 22,* 1183–1190. http://dx.doi.org/10.1177/0956797611419169

Sanchez-Pages, S., Rodriguez-Ruiz, C., & Turiegano, E. (2014). Facial masculinity: How the choice of measurement method enables to detect its influence on behaviour. *PLoS ONE, 9,* e112157. http://dx.doi.org/10.1371/journal.pone.0112157

Saribay, S. A., Biten, A. F., Meral, E. O., Aldan, P., Třebický, V., & Kleisner, K. (2018). The Bogazici face database: Standardized photographs of Turkish faces with supporting materials. *PLoS ONE, 13,* e0192018. http://dx.doi.org/10.1371/journal.pone.0192018

Schaefer, K., Mitteroecker, P., Fink, B., & Bookstein, F. L. (2009). Psychomorphospace—From biology to perception, and back: Towards an integrated quantification of facial form variation. *Biological Theory, 4,* 98–106. http://dx.doi.org/10.1162/biot.2009.4.1.98

Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22,* 32–70. http://dx.doi.org/10.1177/1088868317698288

Scott, I. M. L., Pound, N., Stephen, I. D., Clark, A. P., & Penton-Voak, I. S. (2010). Does masculinity matter? The contribution of masculine face shape to male attractiveness in humans. *PLoS ONE, 5,* e13585. http://dx.doi.org/10.1371/journal.pone.0013585

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15,* 72–101. http://dx.doi.org/10.2307/1412159

Stern, C., & Rule, N. O. (2018). Physical androgyny and categorization difficulty shape political conservatives' attitudes toward transgender people. *Social Psychological and Personality Science, 9,* 24–31. http://dx.doi.org/10.1177/1948550617703172

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127,* 105–118. http://dx.doi.org/10.1016/j.cognition.2012.12.001

Taga, F. (2005). East Asian masculinities. In M. S. Kimmel, J. Hearn, & R. W. Connell (Eds.), *Handbook of studies on men and masculinities* (pp. 129–140). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781452233833

Tan, A. (2016). *Merriam-Webster adds more gender-inclusive words to dictionary*. ABC News. https://abcnews.go.com/U.S./merriam-webster-adds-genderqueer-genderfluid-gender-neutral-title/story?id=38675250

Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York, NY: McGraw-Hill. http://dx.doi.org/10.1037/13514-000

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452–460. http://dx.doi.org/10.1016/S1364-6613(99)01403-5

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13,* 724–738. http://dx.doi.org/10.1037/a0032335

Valentova, J., Rieger, G., Havlicek, J., Linsenmeier, J. A. W., & Bailey, J. M. (2011). Judgments of sexual orientation and masculinity-femininity based on thin slices of behavior: A cross-cultural comparison. *Archives of Sexual Behavior, 40,* 1145–1152. http://dx.doi.org/10.1007/s10508-011-9818-1

Watkins, C. D., DeBruine, L. M., Little, A. C., & Jones, B. C. (2012). Social support influences preferences for feminine facial cues in potential social partners. *Experimental Psychology, 59,* 340–347. http://dx.doi.org/10.1027/1618-3169/a000162

Waynforth, D., Delwadia, S., & Camm, M. (2005). The influence of women's mating strategies on preference for masculine facial architecture. *Evolution and Human Behavior, 26,* 409–416. http://dx.doi.org/10.1016/j.evolhumbehav.2005.03.003

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125. http://dx.doi.org/10.1177/01461672992512005

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., . . . Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences, 282,* 20151351. http://dx.doi.org/10.1098/rspb.2015.1351

Widaman, K. F. (2018). On common factor and principal component representations of data: Implications for theory and for confirmatory replications. *Structural Equation Modeling, 25,* 829–847. http://dx.doi.org/10.1080/10705511.2018.1478730

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl, Vol. 1: Matters of public interest; Vol. 2: Personality and psychopathology* (pp. 89–113). Minnesota: University of Minnesota Press.

Wilkinson, S. (2015, March 24). Butch chic: How the gender-neutral trend has ruined my wardrobe. *The Guardian*. https://www.theguardian.com/fashion/2015/mar/24/butch-chic-how-the-gender-neutral-trend-has-ruined-my-wardrobe

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Advances in Experimental Social Psychology, 46,* 55–123. http://dx.doi.org/10.1016/B978-0-12-394281-4.00002-7

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

Zaidi, A. A., White, J. D., Mattern, B. C., Liebowitz, C. R., Puts, D. A., Claes, P., & Shriver, M. D. (2019). Facial masculinity does not appear to be a condition-dependent male ornament and does not reflect MHC heterozygosity in humans. *Proceedings of the National Academy of Sciences of the United States of America, 116,* 1633–1638. http://dx.doi.org/10.1073/pnas.1808659116

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science, 26,* 237–242. http://dx.doi.org/10.1177/0963721416683996

Received May 30, 2020
Revision received August 13, 2020
Accepted August 20, 2020 ■