Gender/Sex Categories and Gendered Cues in People Perception: The Influence of Gender/Sex Ratio and Gendered Appearance on Group Judgments

Nicholas P. Alt[1]

Pamala N. Dayley[2]

Anna G. Faulkner[1]

Kerri L. Johnson[2,3]

[1] Department of Psychology, Occidental College

[2] Department of Psychology, University of California, Los Angeles

[3] Department of Communications, University of California, Los Angeles

Author Note

Nicholas P. Alt https://orcid.org/0000-0001-6186-6463

Pamala N. Dayley https://orcid.org/0000-0001-8955-9502

Kerri L. Johnson https://orcid.org/0000-0002-1458-2019

We have no known conflict of interest to disclose.

This submission is a: **Research Report/Research Article**

Correspondence concerning this article should be addressed to Nicholas P. Alt, Department of Psychology, 1600 Campus Rd., Swan Hall 336, Los Angeles, CA 90041. Email: alt@oxy.edu This research was supported by a National Science Foundation grant (BCS 2017245/BCS 2017250) awarded to N. P. Alt and K. L. Johnson.

Word Count: 7,821

Table and Figures: 1

Abstract

Previous findings on people perception show perceivers are attuned to the social categories of group members which subsequently influences social judgments. An outstanding question is whether perceivers are also attuned to visual cue variability (e.g., gender typicality). In two studies (n = 165) perceivers viewed twelve-person ensembles (500 ms) of varying White men-to-women ratios. Importantly, faces of one gender/sex were morphed to appear either more masculine or feminine. Consistent with prior work, results indicated judgments varied by the actual gender/sex ratio. Additionally, perceivers' judgments varied as a function of manipulated gender cues. Ensembles composed of masculine, compared to feminine White men, were judged to have more men, higher perceived masculinity, and to be more threatening. Complimentary results were found for ensembles composed of feminine, compared to masculine White women. These findings highlight the impact of both social categories and visual phenotypic cue variability on people perception.

Keywords: people perception, social vision, ensemble coding, gender/sex

Gender/Sex Categories and Gendered Cues in People Perception: The Influence of Gender/Sex Ratio and Gendered Appearance on Group Judgments

When we see a group, be it at a festival, a bar, a religious ceremony, or any other daily occurrence, we form a general impression or gist of who comprises that group. We might consider whether the group is composed of men and women, White and non-White individuals, old or young people. Furthermore, within these social categories, there exists variability in the visual cues that each person embodies. For instance, a man's facial features might be more or less masculine; a young person might look more or less mature than their actual age; and a person's racial category membership might be clear or more ambiguous. This phenotypic variability in visual cues is vital to both social categorization and evaluative processes of faces judged in isolation (Johnson et al., 2015; Maddox, 2004; Oh et al., 2018; Zebrowitz & Montepare, 1992). Little is known, however, about how this variability impacts judgments of groups of faces. Here, we extend research on people perception (Alt & Phillips, 2022), that is how perceivers interpret groups of people via ensemble perception processes (Whitney & Yamanashi Leib, 2018), to examine whether perceivers are sensitive not only to the relative ratios of men and women (e.g., social categories) within an ensemble but also the gendered appearance of the men and women (e.g., phenotypic variability in visual cues). Based on social vision research, we propose that perceivers are likely to integrate both social category and visual cue information when making judgments about ensembles. We tested this by showing participants ensembles of faces that varied not only in the ratio of men to women, but also in the phenotypic masculinity or femininity of men (Study 1) and women (Study 2) to examine the

extent to which perceived gender/sex categories and perceived gender cues[1] inform perceptual judgments and threat evaluations of ensembles.

Ensemble Coding Research on Social Categories

The burgeoning area of ensemble perception (Whitney & Yamanashi Leib, 2018) has revealed that our perceptual systems are attuned to statistical properties (e.g., average and variance) extracted from our sensory experiences. This sensitivity has been demonstrated for a variety of stimuli ranging from low-level features such as the orientation of a group of lines (Dakin & Watt, 1997) or the average size of a set of circles (Ariely, 2001) to more social stimuli such as the average emotional expression of a group (Haberman & Whitney, 2007), the distribution of social categories (Alt et al., 2019; Goodale et al., 2018), and the appearance of dominance and trustworthiness of faces (Chwe & Freeman, 2023; Phillips et al., 2018). Beyond just averaging, judgments related to numerosity, that is the ability to report the number of targets representing a particular social category such as race or gender/sex, are also accurate (Yang & Dunham, 2019). Overall, ensemble perception processes allow our perceptual systems to simplify complex stimuli, relying on a different set of mechanisms (e.g., averaging) to efficiently handle the expanded capacity demands (i.e., ensembles being composed of multiple individuals compared to a single face) and the inherent complexity of group percepts (Alt & Phillips, 2022; Phillips et al., 2014).

While much of the early work on ensemble perception focused on perceivers' accuracy, social psychologists have begun to examine how perceiving ensembles informs our assessment

[1] We use the term gender/sex to align with theorizing about the intertwined nature of these categories (Hyde et al., 2019; van Anders, 2015). We explicitly note that gender/sex should not be strictly considered binary and recognize that our use of language and conceptualization of our methods does not fully encompass the diversity of gender/sex categories (see Hyde et al., 2019; Morgenroth & Ryan, 2021). To our knowledge, all stimuli used in this study were photos of people who identified as cis-gendered men or women (see Ma et al., 2015).

of emergent group properties. For instance, Alt and colleagues (2019) showed that perceivers' judgments of the number of men within a group inform judgments of threat, such that when the ratio of men to women increases, groups are judged to be more threatening. In addition, perceivers' own gender identity interacts with a group's gender/sex ratio to influence feelings of fit and belonging (Goodale et al., 2018). Ensemble-based features can also affect whether we categorize other groups as our group or not. Specifically, Goldenberg and colleagues (2020) used a minimal groups paradigm to create two artificial groups. They then showed participants groups that had high or low emotion expression variance to evocative stimuli (e.g., immoral acts). Participants were asked whether the viewed group was their group or not and found that participants self-identified with the low variance groups over the high variance groups. Thus, group emotion variance predicted judgments of group categorization. Other researchers have shown that features such as group eye gaze can influence trust ratings as perceivers gave higher trustworthiness ratings to groups when the groups' eye gaze signaled where a subsequently presented target would appear, compared to groups where eye gaze was not a valid signal (Sun et al., 2020). The very act of seeing a group can also compel shifts in intergroup attitudes. Participants who viewed interracial ensembles that depicted high, relative to low, emotional segregation reported higher levels of race essentialism (Lamer et al., 2018). Collectively, this research showcases that perceivers are attuned to visual group properties and that they use this information as the basis for social evaluations.

Social Vision and the Influence of Bottom-Up Visual Cues

While much of the work above focuses on specific social categories (e.g., gender/sex) or holistic impressions of traits (e.g., trust), it is critical to note that the visual cues that underlie these social categories (e.g., gendered appearance) vary and that this variability is consequential.

The importance of variability in visible cues is a central tenet of the emerging body of work probing social vision processes, which explicitly takes into account both visible cues that originate in the target of perception and also existing knowledge structures that originate in the perceiver (e.g., stereotypes and motives) to elucidate how observers make social categorical and evaluative judgments for visual stimuli (Adams et al., 2010; Balcetis & Lassiter, 2010; Freeman & Ambady, 2011; Johnson et al., 2015; Oswald & Adams, 2022). For instance, in the domain of gender/sex, visual cues to masculinity and femininity vary across men and women, and this variability influences downstream judgments (Johnson & Adams, 2013; Oosterhof & Todorov, 2008). For example, Friedman and Zebrowitz (1992) manipulated men's and women's faces to appear more or less "baby-faced" (e.g., by increasing the size of the eyes and the forehead-to-chin ratio). This change in visual cues influenced gendered trait ratings such that baby-faced men were viewed as warmer and more caring than their non-baby-faced counterparts and baby-faced women were viewed as less dominant than their non-baby-faced counterparts. Gendered facial appearance is also associated with judgments of dominance and leadership, with more masculine targets (both men and women) rated higher on these traits than more feminine targets (Sczesny et al., 2006).

An interesting question for people perception emerges when considering the influence of visual cue variability and social category differences in ensembles. While perceivers readily discern social categories within an ensemble (e.g., men and women, Alt et al., 2019; Goodale et al., 2018; Yang & Dunham, 2019), it is less clear whether this is due to social categorization (i.e., number of men and women), an averaging of visual cue variability (i.e., more masculine/feminine targets), or a combination. Essentially, different routes could lead to the same people perception outcomes (e.g., group categorization). One perspective rooted in

classical social categorization theory, suggests that since perceivers spontaneously categorize others (Allport, 1954; Brewer, 1988; Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000) this attunement to solely the social category may lead to accurate ensemble perception with limited influence of phenotypic variability (e.g., gendered appearance of the faces in the ensemble). From another perspective rooted in social vision findings such as how social categorization decisions unfold in real-time (Freeman et al., 2008), we know that phenotypic cues are processed dynamically and can influence how a social category decision is reached. Applied to people perception, summary percepts might also operate via a dynamic integration of phenotypic variability. Thus, judgments of ensembles may vary systematically due to visual cue variability (e.g., whether the men and women in the ensemble are more masculine or feminine). Perhaps most likely, both processes are operating at the same time, with information from social category and phenotypic variability influencing judgments of social ensembles, such that perceivers are sensitive to the number of social category members within an ensemble and also incorporate the visual cue variability that exists. These studies seek to test each of these possibilities.

Notably, some prior work on ensemble perception has, by virtue of the methods employed, examined sensitivity to visual cues of gender/sex; yet such work has not probed the social-evaluative implications of gender appearance information. For example, Haberman and Whitney (2007) showed ensembles of four faces that varied in gendered cues and assessed the accuracy of average ensemble gender appearance judgments. The ensembles, however, were composed of two facial identities (one man and one woman) that were morphed along a continuum. While important for establishing that perceivers are sensitive to gendered cues, this is not reflective of our day-to-day visual experience insofar as it would be odd to perceive a group

composed of just two facial identities that varied in terms of their degree of masculinity and femininity. In addition, Phillips and colleagues (2018) showed that perceivers were sensitive to gendered appearance variability within four-person ensembles (i.e., is this ensemble more or less variable in terms of gendered appearance than a previously viewed ensemble). While the number of identities used was increased to ten, the task focused on perceptual ability and not the downstream social-evaluative consequences of perceiving ensembles that vary in both gender/sex ratio and gendered appearance. Nevertheless, these studies provide initial evidence that perceivers are sensitive to gendered cue variability within ensembles. When combined with social vision research on single faces, existing research suggests that observed variability in gendered cues should also affect downstream evaluative judgments of ensembles.

Present Research

In the current studies, we tested whether ensemble-based judgments are informed by both the social categories (i.e., gender/sex) within the ensemble, and also visible variation in phenotypic cues of the individuals within the ensemble (i.e., variation in gendered appearance). Aligning with prior research (Alt et al., 2019), we hypothesized that participants' numeric estimations of the number of men (Study 1) and women (Study 2) and perceived average gender appearance would be positively associated with the actual number of men and women in the ensemble and that threat judgments would increase with the number of men, or conversely decrease with the number of women, in the ensemble. Critically, we also hypothesized a significant interaction between the actual ratio of men to women within an ensemble and the manipulated gendered appearance of the targets within the ensemble. Specifically, we predicted that when ensembles depicted faces of men that were manipulated to appear more masculine, compared to feminine, ensembles would elicit accentuated judgments for the perceived number

of men, perceived masculinity of the group, and perceived threat. Conversely, we predicted that when ensembles depicted faces of women that were manipulated to appear more feminine, compared to masculine, ensembles would elicit accentuated judgments for the perceived number of women, perceived femininity of the group, and attenuated judgments of perceived threat. We tested these predictions in two studies, one that manipulated the gender appearance of the men within each ensemble (Study 1) and one that manipulated the gender appearance of the women within each ensemble (Study 2).

Study 1

Method

Participants

We recruited 96 participants (22 men, 74 women, M_{age} = 19.17, SD_{age} = 0.99, 30.2% Asian/Asian American, 21.9% White/European American) through the UCLA research subject pool. Sample size was based on prior people perception studies (see Alt et al., 2019; Goodale et al., 2018) and the rate of data collection in the lab. Given the issues of conducting power analyses for multilevel models (e.g., requiring information about covariance structure), post hoc sensitivity analyses were run using the simr package (Green & MacLeod, 2016) in R. Results indicated we had a sufficient number of participants to detect the effects found within our results with 99% power. All participants were given course credit for participating in the study.

Stimuli

We used Abrosoft FantaMorph 5.4.8 (Abrosoft, 2020) facial morphing software to create a set of men's faces that differed in terms of gendered appearance. First, we created composite facial images of men and women that served as anchors to manipulate gendered appearance. To create these anchors, we randomly selected five White men’s and five White women’s faces

(without replacement) from the Chicago Face Database (Ma et al., 2015). Using FantaMorph, we adjusted 114 facial landmarks on each face and morphed the selected faces into a composite man's and woman's face. This process was repeated until we had ten composite facial images of men and ten composite facial images of women.

Next, we used these composite faces to manipulate the gendered appearance of men's faces to appear either more masculine (i.e., morphed to a composite man's face) or more feminine (i.e., morphed to a composite woman's face). Using the same morphing procedure described above, we morphed sixty-six White men's faces from the Chicago Face Database (excluding faces that were categorized below 80% White based on pre-test data, faces with hair below the shoulders, and faces that produced artifacts in the morphing, see Ma et al., 2015), 35% towards a composite man or composite woman face. Composite faces were randomly selected until we needed to repeat the use of composite faces (i.e., every ten morphs given we had ten composite faces). To account for subtle changes created by face morphing (e.g., smoother skin, hair artifacts, see Sutherland et al., 2017) we also morphed White women's faces, selected from the Chicago Face Database (n = 72), 35% to a randomly selected woman's composite face. Overall, our stimuli set from which groups were created (see below) consisted of 66 White masculine men's faces, 66 White feminine men's faces, and 72 White women's faces (that were morphed to a composite woman's face)[2].

Similar to past work (Alt et al., 2019), each ensemble consisted of a 3x4 visual array composed of men's and women's faces. A Python script generated ensembles, in real-time, with

[2] We conducted a pilot test on a subset of the faces (n = 36, 18 masculine men, 18 feminine men) to ensure that participants viewed the morphs as differentiated based on gendered appearance. Twenty participants recruited online rated each face for gender appearance (1 = very masculine, 7 = very feminine). Results indicated that the masculinized faces, on average, were rated as significantly more masculine (M = 2.20, SD = 0.56) than the femininized faces (M = 2.73, SD = 0.73), $t(34) = -4.62$, $p < .001$, 95% CIs [-1.45, -0.56].

random selection and random placement of the faces within a 3x4 array (Goodale et al., 2018). Faces were selected from the available pool of faces with the limitation that each ensemble was composed of a unique set of faces (i.e., no face appeared more than once within each ensemble). In total, 150 ensembles were shown to participants with 15 ensembles composed of masculine men's faces and 15 ensembles composed of feminine men's faces for each of the following ratios: 0 men:12 women (30 trials of only women's faces)[3], 3 men:9 women, 6 men:6 women, 9 men:3 women, and 12 men:0 women.

Procedure

Each trial consisted of a fixation cross for 500 ms (non-overlapping with the location of any face to appear in the subsequent ensemble), followed by the ensemble for 500 ms, and then the prompt for a judgment. This procedure was repeated across three counterbalanced blocks. In one block, participants estimated the Perceived Number of Men in each ensemble (*0 men to 12 men*). In another block, participants rated each ensemble on four evaluative dimensions of Perceived Threat: invitingness (1 = *not at all inviting,* 7 = *extremely inviting,* reverse coded), warmth (1 = *extremely cold,* 7 = *extremely warm,* reverse coded), hostility (1 = *not at all hostile,* 7 = *extremely hostile*), and threat (1 = *not at all threatening,* 7 = *extremely threatening*). Finally, participants judged the Perceived Average Gender Appearance of the ensemble using a visual gender continuum scale created from FaceGen faces. This scale depicted 21 faces that varied systematically from hyper-feminine to hyper-masculine. Participants indicated via numbers under each face (i.e., 1-21) which face best represented the average gender appearance of the group (i.e., "Please select the face which best represents the mean or average facial

[3] Since these trials contained no faces of men, we randomly coded half the trials as masculine and half the trials as feminine. If we exclude these trials from analyses, our results remain consistent.

masculinity/femininity from the group"). After all blocks were completed, participants provided demographic information and were debriefed.

Results

We hypothesized that judgments of the Perceived Number of Men, Perceived Average Gender Appearance and Perceived Threat would increase as the actual ratio of men to women in each ensemble increased, replicating and extending prior findings. This pattern would corroborate the role of social categorization in ensemble percepts insofar as it establishes a perceptual sensitivity to the actuarial distribution of men to women in an ensemble. More importantly, we also predicted that these patterns would be moderated by Manipulated Gendered Appearance, such that for ensembles composed of masculine, compared to feminine men, the rate of increase (i.e., slope) for each of these measures would be steeper. This pattern would establish for the first time that observers are sensitive not only to the categorical distribution of social categories within an ensemble, but also to the phenotypic variations exhibited within a subset of the ensemble's members, here men.

To test this, we used hierarchical linear models with fixed effects for our manipulated variables Actual Number of Men (mean-centered) and Manipulated Gendered Appearance (i.e., whether the men's faces were morphed toward masculine or feminine anchors (effect coded, 1 = feminine, -1 = masculine), and the interaction between the two. Models included a random factor for participants with random intercepts and random slopes for the Actual Number of Men. We take this approach to account for the repeated measures structure of the data and participants' idiosyncratic responses (see Judd et al., 2016 for more information). Analyses were conducted in R using lmer and lmerTests (Bates et al., 2015; Kuznetsova et al., 2015). For all analyses, we first included Participant Gender as a factor in each analysis. For Study 1, because we found no

significant interactions involving Participant Gender, it will receive no further mention. All data and analysis code for Studies 1 and 2 can be found at https://osf.io/vn4tk/?view_only=2858d3bf203f4563b1f5508c05398f54.

Perceived Number of Men

As the Actual Number of Men increased, so did participants' Perceived Number of Men, B = 2.54, SE = 0.08, t = 33.27, p < .001, 95% CIs [2.39, 2.69]. Manipulated Gendered Appearance was also significant, with ensembles depicting feminine, compared to masculine, men eliciting lower estimates for the Perceived Number of Men, B = -0.37, SE = 0.02, t =-18.40, p < .001, 95% CIs [-0.41, -0.33]. Importantly, there was a significant interaction between Actual Number of Men and Manipulated Gendered Appearance (see Fig. 1), B = -0.30, SE = 0.02, t = -15.24, p < .001, 95% CIs [-0.34, -0.27]. Tests of simple slopes revealed that the relation between Actual Number of Men and Perceived Number of Men was steeper when men's faces were morphed to appear more masculine, B = 2.85, SE = 0.08, t = 36.06, p < .001, 95% CIs [2.69, 3.00], than when they were manipulated to appear more feminine, B = 2.24, SE = 0.08, t = 28.32, p < .001, 95% CIs [2.08, 2.40].

Perceived Average Gender Appearance

A similar pattern of findings emerged for Perceived Average Gender Appearance. As the Actual Number of Men increased, participants judged the Perceived Average Gender Appearance to be more masculine, B = 2.75, SE = 0.13, t = 20.78, p < .001, 95% CIs [2.49, 3.01]. In addition, there was an overall effect of Manipulated Gendered Appearance, such that when men's faces were morphed to appear feminine, relative to masculine, ensembles were also judged to be less masculine, overall, B = -0.40, SE = 0.03, t = -13.15, p < .001, 95% CIs [-0.46, -0.34]. Again, a significant interaction between the Actual Number of Men and Manipulated

Gendered Appearance emerged, B = -0.32, SE = 0.03, t = -10.51, p < .001, 95% CIs [-0.38, -0.26]. Tests of simple slopes indicated that when men's faces were morphed to appear more masculine there was a steeper slope for Actual Number of Men, B = 3.07, SE = 0.14, t = 22.59, p < .001, 95% CIs [2.80, 3.33] compared to when men's faces were morphed to appear more feminine, B = 2.43, SE = 0.14, t = 17.91, p < .001, 95% CIs [2.16, 2.69].

Perceived Threat

A mean score was computed from the four evaluative items, with warmth and invitingness reverse coded, and the within-subject reliability was .98 (see Cranford et al., 2006 for method). As the Actual Number of Men in the ensemble increased, participants' judgments for Perceived Threat increased, B = 0.32, SE = 0.04, t = 8.47, p < .001, 95% CIs [0.25, 0.40]. There was also an effect of Manipulated Gendered Appearance, such that ensembles that depicted feminine, relative to masculine, men's faces were judged to be less threatening, B = -0.06, SE = 0.01, t = -7.97, p < .001, 95% CIs [-0.08, -0.05]. These results were qualified by a significant interaction between the Actual Number of Men and Manipulated Gendered Appearance (see Fig. 1), B = -0.05, SE = 0.01, t = -6.81, p < .001, 95% CIs [-0.07, -0.04]. Tests of simple slopes indicated that when the men's faces were morphed to appear masculine, there was a steeper slope for Actual Number of Men, B = 0.38, SE = 0.04, t = 9.69, p < .001, 95% CIs [0.30, 0.46] compared to when the men's faces were morphed to appear feminine, B = 0.27, SE = 0.04, t = 6.88, p < .001, 95% CIs [0.19, 0.35].

Study 1 Discussion

Overall, our results show that both the actual ratio of men to women and variations in visible cues to gender/sex influenced judgments of the perceived number of men in the ensemble, the perceived average gender appearance, and the perceived threat of the ensemble.

These patterns replicate prior work showing that perceivers are sensitive to the distribution of social categories within ensembles and that they render social-evaluative judgments in response to categorical variability across ensembles (Alt et al., 2019). Additionally, these findings demonstrate that perceivers are sensitive to variations in visible cues *within* a social category, here systematic variations in the gendered appearance of the men depicted in each ensemble. As such, these findings provide important insights into the mechanisms of ensemble perception. Indeed, observers attended to both categorical and phenotypic variation when rendering judgments of numerosity, perceived average gender appearance, and perceived threat.

The findings of Study 1 therefore provide important information that expands our understanding of ensemble perception, yet they did so only for variations in the appearance of men. Because current society remains andro-centric (Bailey et al., 2019; Stroessner, 1996), observers may be sensitive to variations in the appearance of men, but not of women. In Study 2, we conceptually replicated these findings, this time manipulating the appearance of women's faces within each ensemble to appear either more masculine or more feminine. We hypothesize that again perceivers' judgments would be affected by both social category information (e.g., the number of women within the group) and the manipulation of visible gendered cues of the women depicted within each ensemble (e.g., whether the women are masculine or feminine).

Study 2

Method

Participants

We recruited 69 participants (14 men, 51 women, 3 non-binary, M_{age} = 18.94, SD_{age} = 1.92, 42.0% White/European American, 18.8% Asian/Asian American, 15.9% Latinx American) from the UCLA research subject pool. Same as Study 1, post hoc sensitivity analyses were

conducted and results indicated we had a sufficient number of participants to detect the effects found within our results with 99% power. Participants received course credit for participating in the study.

Stimuli

We utilized the same stimuli generation process described in Study 1 to create a set of women's faces that were morphed to be either more masculine or feminine. We again used FantaMorph and the 10 composite men's faces and women's faces used from Study 1 but this time morphed White women's faces from the Chicago Face Database (Ma et al., 2015). We morphed seventy-two White women's faces from the Chicago Face Database (excluding faces that were categorized below 80% White based on pre-test data and faces that produced artifacts in the morphing, see Ma et al., 2015), 35% towards a composite man or woman face (randomized but repeated after all 10 composite faces were used). The men's faces used in ensembles were the men's faces that were morphed to be masculine from Study 1. Overall, the stimuli set from which ensembles were created consisted of 72 White masculine women's faces, 72 White feminine women's faces, and 66 White men's faces.

Using the same process as Study 1, ensembles were created by randomly selecting and randomly placing faces within a 3x4 array. In total, 150 ensembles were shown to participants with 15 ensembles composed of masculine women's faces and 15 ensembles composed of feminine women's faces for the following ratios: 0 men:12 women, 3 men:9 women, 6 men:6 women, 9 men:3 women, and 12 men:0 women (30 trials of only men' faces)[4]. Similar to Study

[4] Since these trials contained no faces of women, we randomly coded half the trials as masculine and half the trials as feminine as done in Study 1. Results remain consistent if instead, we exclude these trials.

1, participants rated the same set of ensembles across the three blocks (presented in random order within each block).

Procedure

The same procedure paradigm used in Study 1 was followed in Study 2. Across three randomized blocks participants saw a fixation cross for 500 ms, an ensemble for 500ms, and then reported the Perceived Number of Women, Perceived Average Gender Appearance, and Perceived Threat.

Results

Aligning with Study 1, we hypothesized that judgments of the Perceived Number of Women would increase and Perceived Average Gender Appearance and Perceived Threat would decrease as the actual ratio of women to men in each ensemble increased. In addition, we predicted Manipulated Gendered Appearance would moderate judgments such that the slopes for each of these measures would be steeper for ensembles composed of feminine, compared to masculine women. The analytic strategy was the same as Study 1.

Numeric Estimations of Women

Results revealed that as the Actual Number of Women increased, so did participants' Perceived Number of Women, $B = 2.55$, $SE = 0.09$, $t = 29.82$, $p < .001$, 95% CIs [2.38, 2.71]. Manipulated Gendered Appearance was also significant with ensembles depicting feminine women resulting in higher estimates for Perceived Number of Women, $B = 0.36$, $SE = 0.02$, $t = 16.16$, $p < .001$, 95% CIs [0.31, 0.40]. Importantly, there was a significant interaction between Actual Number of Women and Manipulated Gendered Appearance (see Fig. 1), $B = 0.27$, $SE = 0.02$, $t = 12.35$, $p < .001$, 95% CIs [0.23, 0.32]. Tests of simple slopes revealed that the relation between Actual Number of Women and Perceived Number of Women was steeper when

women's faces were morphed to appear more feminine, $B = 2.82$, $SE = 0.09$, $t = 31.96$, $p < .001$, 95% CIs [2.64, 2.99], than when they were morphed to appear more masculine, $B = 2.27$, $SE = 0.09$, $t = 25.78$, $p < .001$, 95% CIs [2.10, 2.45].

Perceived Average Gender Appearance

Results also aligned with hypotheses for Perceived Average Gender Appearance. As the Actual Number of Women increased, participants judged the Perceived Average Gender Appearance of the ensemble to be less masculine, $B = -3.00$, $SE = 0.21$, $t = -14.56$, $p < .001$, 95% CIs [-3.40, -2.59]. In addition, there was an overall effect of Manipulated Gendered Appearance, such that when women's faces were morphed to appear feminine, relative to masculine, ensembles were judged to be on average less masculine, $B = -0.37$, $SE = 0.04$, $t = -10.37$, $p < .001$, 95% CIs [-0.44, -0.30]. Again, a significant interaction between the Actual Number of Women and Manipulated Gendered Appearance emerged, $B = -0.27$, $SE = 0.04$, $t = -7.66$, $p < .001$, 95% CIs [-0.34, -0.20]. Tests of simple slopes indicated that when women's faces were morphed to appear feminine there was a steeper slope for Actual Number of Women, $B = -3.27$, $SE = 0.21$, $t = -15.65$, $p < .001$, 95% CIs [-3.68, -2.86] compared to when the women's faces were morphed to appear masculine, $B = -2.73$, $SE = 0.21$, $t = -13.05$, $p < .001$, 95% CIs [-3.14, -2.32].

Perceived Threat

A mean score was computed from the four evaluative items, with warmth and invitingness reverse coded, and the within-subject reliability was .94 (see Cranford et al., 2006 for method). As the Actual Number of Women in the group increased, participants' judgments for Perceived Threat decreased, $B = -0.44$, $SE = 0.05$, $t = -8.97$, $p < .001$, 95% CIs [-0.53, -0.34]. There was also an effect of Manipulated Gendered Appearance, such that ensembles that

depicted feminine, compared to masculine, women's faces were judged less threatening, (see Fig. 1), B = -0.11, SE = 0.01, t = -11.48, p < .001, 95% CIs [-0.13, -0.09]. These results were qualified by a significant interaction between the Actual Number of Women and Manipulated Gendered Appearance, B = -0.10, SE = 0.01, t = -10.51, p < .001, 95% CIs [-0.12, -0.08].[5] Tests of simple slopes indicated that when the women's faces were morphed to appear feminine there was a steeper slope for Actual Number of Women, B = -0.71, SE = 0.06, t = -11.32, p < .001, 95% CIs [-0.83, -0.59] compared to when the women's faces were morphed to appear more masculine, B = -0.50, SE = 0.06, t = -8.05, p < .001, 95% CIs [-0.63, -0.38].

Study 2 Discussion

Study 2 results replicated Study 1 however using women's faces that were morphed to be masculine or feminine. We again saw that perceivers were attuned to both social category (i.e., the number of women to men) and visual phenotypic variability (i.e., women's gender typicality). Specifically, as the number of women in the ensemble increased, participants reported seeing more women, that the perceived average gender appearance of the group was more feminine, and that perceived threat decreased. Furthermore, these effects were stronger for ensembles composed of feminine women compared to masculine women. Together with Study 1,

[5] This was the only case where we found a significant three-way interaction with Participant Gender (coded as -1 = men, 1 = women), B = 0.03, SE = 0.01, t = 2.46, p = .014, 95% CIs [0.01, 0.05]. Breaking down this three-way interaction by Participant Gender, we found that our two-way interaction of interest (i.e., Actual Number of Women by Manipulated Gendered Appearance) was not significant for men participants, B = -0.05, SE = 0.03, t = -1.64, p = .102, 95% CIs [-0.11, 0.01], but was significant for women participants, B = -0.12, SE = 0.02, t = -7.86, p < .001, 95% CIs [-0.15, -0.09]. Simple effect tests showed that for women participants, when women's faces were morphed to appear feminine there was a steeper slope for Actual Number of Women, B = -0.81, SE = 0.07, t = -12.32, p < .001 95% CIs [-0.94, -0.68] compared to when women's faces were morphed to appear more masculine, B = -0.57, SE = 0.07, t = -8.66, p < .001 95% CIs [-0.70, -0.44]. That being said, we caution against a strong interpretation of this three-way interaction because we did not set out to recruit equal numbers of men and women participants and indeed had significantly more women (n = 51) compared to men (n = 15).

these findings strongly suggest that ensemble perception processes are attuned not just to social category differences within ensembles but also to the phenotypic variability within visual cues of the ensemble.

General Discussion

As ensemble perception research becomes more common in social psychological research, generally, and social vision research, specifically, several key insights have been established. First, perceivers are attuned to social categories such as gender/sex and race within groups; from merely a glimpse, perceivers accurately report the ratio or number of social categories members depicted (Alt et al., 2019; Phillips et al., 2018; Yang & Dunham, 2019). Moreover, these groups affect social-evaluative judgments such as perceived threat and feelings of fit and belonging (Alt et al., 2019; Goodale et al., 2018). Although this prior work provided evidence that perceivers accurately detect social categories within groups, it provided no insight into whether and how variability in the phenotypic visual cues of the target of perception (e.g., appearing more or less masculine) might also be perceived. This left open the question of whether observers achieve accurate percepts *solely* from numeric accounting of the distribution of different social categories (e.g., this group is majority women) or whether they also incorporate variability in the visual cues to social categories (e.g., this group is majority feminine women). The current studies shed light on this issue.

In two studies, groups were composed of men's and women's faces that were masculinized and femininized using facial morphing. We found that perceivers were sensitive not only to the social categories represented within ensembles (i.e., number of women and men) but also the variability in visual cues that signaled those social categories (i.e., masculinity and femininity). We consistently found an interaction between a group's actual gender/sex ratio and

manipulated gendered appearance for a range of social perception judgments. Specifically, when ensembles depicted either masculinized (relative to feminized) men or feminized (relative to masculinized) women, participants reported: a) higher numeric estimates for the number of men (Study 1) and women (Study 2) in each ensemble, b) that the perceived average gender appearance of the group was more masculine (Study 1) and feminine (Study 2), and c) that the group was more (Study 1) or less (Study 2) threatening. This pattern of results reveals that people perception processes operate not only at the social category level but also at the visual phenotypic level, affecting actuarial estimates, perceptions of the group's average gendered appearance, and social-evaluative judgments.

These findings provide an important link between ensemble perception research within vision sciences (Whitney & Yamanashi Leib, 2018) and social vision research (Adams et al., 2010; Johnson et al., 2015), revealing new insights about how variability in phenotypic cues affect people perception. Early vision science work on ensemble perception largely focused on determining whether perceivers were attuned to visual cues, with a strong emphasis on determining accuracy. For instance, Haberman and Whitney (2007) found that perceivers could estimate the average "gender" of a group of four faces, which varied in only two identities. Yet their methods were decidedly within vision science (n = 2, offset by a large number of trials) and focused squarely on accuracy, leaving open questions about the broader social implications of ensemble perception. The current work leveraged observations that phenotypic variability affects both social categorizations and evaluations of individual targets (Freeman et al., 2008; Johnson & Tassinary, 2005; Ward et al., 2012) to extend our understanding of the perception of groups. Combining these social vision insights with vision science methods provides important advances for both fields. Specifically, even at relatively fast presentation times (500 ms), perceivers'

actuarial and social-evaluative judgments are influenced not only by the gender/sex categories represented in each group but also by targets' gendered appearance. Our findings therefore continue to build insights at the nexus of vision science and social psychology that could not be achieved from either alone. This remains a fertile area for continued interdisciplinary research to build our understanding of people perception (Alt & Phillips, 2022).

While the pattern of results across the two studies clearly indicates that both social category representation and gendered appearance combine to affect judgments, there are some questions about the exact mechanism underlying these findings. We have interpreted these results to indicate that perceivers made social categorizations of men and women and integrated gendered appearance variation into their judgments. This possibility entails that perceivers accurately categorize each target (e.g., categorizing a feminine man as a man), but then allowed gendered appearance to upwardly or downwardly bias judgments. However, another possibility is that our manipulation of gendered appearance might have led some targets within the group to be misgendered (e.g., categorizing a feminine man as a woman). While this distinction is difficult to definitively determine given our methods, some patterns in the data are more consistent with the former, rather than the latter, interpretation. Specifically, the overall effect of the actual gender/sex ratio of groups reliably influenced each judgment, indicating that perceivers were attuned to actuarial differences in the number of men and women. These effects were then moderated by our manipulated gendered appearance. Indeed, we found small but reliable shifts in participants' estimated number of men and women as a function of manipulated gendered appearance. This pattern would be unlikely if perceivers had routinely misgendered the masculine women and feminine men in each ensemble. Instead, if misgendering had been prevalent, it should have led to a more dramatic shift in the observed pattern of results, if not a

complete reversal (e.g., ensembles of 12 feminine men and zero women being perceived as having no men). Our findings on perceived threat judgments also corroborate our interpretation of the results. A post-hoc test revealed that groups consisting of twelve feminine men were perceived as more threatening (M = 3.98, SD = 0.75) than groups consisting of twelve masculine women (M = 3.52, SD = 1.17), B = 0.23, SE = 0.05, t = 4.79, $p < .001$, 95% CIs [0.14, 0.32]. This pattern of results is therefore more consistent with gendered appearance biasing perceivers' assessments of social categories, rather than errors in gender/sex categorization.

Second, we used facial morphing software to manipulate the gendered appearance of our stimuli, and this technique is known to smooth blemishes/texture, remove skin imperfections, and increase perceived attractiveness (Kemelmacher-Shlizerman et al., 2014; Langlois & Roggman, 1990; Sutherland et al., 2017). Although *all* faces within each group were morphed (e.g., for groups where we manipulated women's gendered appearance, we also used faces in which men's faces were morphed to other men), there may still be influences of the morphing on judgments. For instance, masculinity may be reduced by the smoothing of the skin and rounding of the jaw (Sutherland et al., 2017), yet this influence would likely *reduce* the magnitude of our effects. Even though morphing tends to minimize cues to masculinity, we nevertheless observed hypothesis-consistent effects. Still, it may be the case that the difference in judgments due to manipulated gendered appearance could be driven by seeing feminine men in Study 1 and feminine women in Study 2, and not an equal contribution of masculine men/women targets as the morphing process makes these targets not as masculine. Future research may wish to examine gendered appearance cues using unmanipulated faces that are pre-rated as high or low on masculinity/femininity to address this issue, however, we anticipate similar, if not stronger, results would emerge.

Our findings also provide a strong foundation to speculate how other phenotypic cues beyond gendered facial appearance, are likely to impact observers' actuarial and evaluative judgments of groups. For instance, within the domain of gender/sex, gendered cues in people's attire are also likely to impact how a group is perceived (Hester & Hehman, 2023) and existing evidence suggests clothing cues are perceptible from groups (Parkinson et al., 2023). In addition, it is well established that phenotypic variations in other perceptible social categories such as race, have profound impacts on both social-evaluative judgments (Blair et al., 2004; Maddox, 2004) and real-world outcomes (Eberhardt et al., 2006). It is therefore likely that judgments of groups composed of different racial categories will also be sensitive to variations in racial phenotypic cues.

These results also have several implications for understanding people perception as groups become more complex. For instance, gender/sex is rarely perceived in isolation, but rather in combination, or intersectionally, with other social categories (Crenshaw, 1989, 1991; Johnson et al., 2015; Kang & Bodenhausen, 2015; Petsko et al., 2022; Remedios & Sanchez, 2018; Sternberg et al., 2023). Of the multiple domains that are likely influenced by gender/sex, we highlight three specific areas. First, and perhaps most notably, gender/sex and race are perceptually intertwined (Goff et al., 2008; Johnson et al., 2012; Lei et al., 2020; Schug et al., 2015; Stroessner, 1996; Zaraté & Smith, 1990). Collectively, this psychological work has revealed that under some circumstances, either race or gender/sex dominates social perception (Stroessner, 1996; Zaraté & Smith, 1990) and that when perceived simultaneously, incidental race perception biases the perception of gender/sex (Johnson et al., 2012), and vice versa (Carpinella et al., 2015). Based on these findings, one might predict that when groups vary simultaneously in both race and gender/sex, perceptual judgments will be biased in a "race is

gendered" fashion, such that the presence of Asian targets will compel more feminine percepts and the presence of Black targets will compel more masculine percepts.

Second, gender/sex and sexual orientation are also perceptually intertwined. Notably, when making judgments about the sexual orientation of individuals, perceivers utilize gender inversion heuristics, which leads gender-typical people to be judged as straight and gender-atypical people to be judged as gay (Freeman et al., 2010; Lick & Johnson, 2014). Thus, same-sex groups of men or women may incidentally elicit sexual orientation judgments as well, such that feminine-appearing groups of men or masculine-appearing groups of women might be perceived as gay and lesbian, respectively. These tendencies are likely to be exacerbated depending on the context in which group perception occurs (e.g., a gay bar versus a sports bar; Lick et al., 2019). Additionally, other visual cues within the space likely inform such judgments (e.g., a pride flag, see Esposito & Calanchini, 2022). Furthermore, as suggested by Alt and Phillips (2022), this attunement to visual cues and social categories within a group may modulate single-person perception judgments. If a self-identified straight man is surrounded by a group of more feminine men, he might be miscategorized as gay via assimilation to the overall group average.

Finally, judgments of emotion are informed by gender/sex (see Adams et al., 2015 for a review) with many of the cues to emotional expression overlapping with gendered appearance (Becker et al., 2007; Hess et al., 2009; Zebrowitz et al., 2010). Such patterns suggest that gendered judgments of groups might also be biased by emotional expressions, such that displays of anger might bias percepts toward male/masculine judgments and displays of either happiness or sadness might bias percepts toward female/feminine judgments. These patterns also extend to body perception (Johnson et al., 2011) inviting the possibility that each of the intersections

described herein might also extend to group perception of *whole* people, rather than of faces alone (Oswald et al., 2023). Collectively, prior work probing intersectional social perception of targets in isolation suggests that people perception might also occur in a combinatorial fashion. These possibilities remain an exciting direction for future research.

Although alluded to throughout this general discussion, some limitations of the current research bear noting explicitly. First, our studies depicted only images of White men and White women. As such, our opining about the possibilities for intersectional people perception above remains speculative, and we cannot speak to how these processes may operate for a more diverse stimulus set, more generally. As one of the first investigations into how phenotypic variability influences people perception, we aimed to minimize complexity and thus limited the diversity of the faces but this remains a constraint future research should address. Our measures were also limited in terms of how we conceptualized masculinity and femininity. Specifically, we used hyper-masculine and hyper-feminine faces as the ends of our perceived average gender appearance scale instead of separating masculinity and femininity. Past work has shown that ratings of masculinity and femininity, from faces, are uniquely associated with trait rating (Hester et al., 2020) suggesting that these two dimensions should not be treated as opposites of the same scale but rather as separate predictors. A future research question would be to examine whether our results hold up if perceived average gender appearance was assessed using separate scales for masculinity and femininity. Notably, however, we did not use the perceived average gender appearance scale as a predictor for other traits but rather as an assessment of the group as a whole. Based on Hester and colleague's (2020) work we would recommend if perceived average gender appearance is measured in relation to other trait ratings of a group, masculinity and femininity should be treated as separate.

We did not explicitly hypothesize, nor did we find consistent effects, that participants' gender moderated our findings. Yet our measurement of participant-level characteristics was limited to gender identity, rather than other aspects of gendered beliefs. It may be the case, for example, that participants' own beliefs about gender (e.g., endorsement of gender essentialism or gender fluidity) may modulate how they perceive and assess gender-linked characteristics within groups (Morgenroth & Ryan, 2020; Oswald & Adams, 2022). Because prior work found that gendered judgments were more strongly influenced by factors in the target of perception than by factors in the perceivers (e.g., how the target looks versus what the perceiver believes, Hehman et al., 2017), studies testing such possibilities will likely require targeted and high-powered tests. Finally, our participant samples consisted of WEIRD populations (Henrich et al., 2010), and greater diversity in participant samples will provide important insights into potential cultural moderation (e.g., holistic versus analytic cognition, Im et al., 2017).

Conclusion

In summary, we leveraged vision science methods and research in social vision to uncover that perceptual and evaluative group judgments incorporate both social categorical information and phenotypic visual cues. Across judgments of the perceived number of men/women, perceived average gender of the group, and perceived threat, participants' ratings were influenced by whether the group included masculinized/feminized, men/women in a manner consistent with hypotheses (e.g., groups containing masculine, compared to feminine, men were rated as more threatening). Our novel insight that both social category (e.g., man and woman) and phenotypic visual cues (e.g., gendered appearance) influence actuarial and social-evaluative judgments of groups provide a foundation for future research on people perception

(Alt & Phillips, 2022) and bridges vision science on ensemble perception and social vision research to better understand how we come to perceive and judge groups.

References

Abrosoft. (2020). *Fantamorph* [Computer software]. https://www.fantamorph.com/

Adams, R. B. A., Jr., Ambady, N., Nakayama, K., & Shimojo, S. (2010). The Science of Social Vision. New York: Oxford University Press.

Adams, R. B., Hess, U., & Kleck, R. E. (2015). The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Review*, *7*(1), 5–13. https://doi.org/10.1177/1754073914544407

Allport, G. W. (1954). The Nature of Prejudice (Unabridged, 25th anniversary ed). Reading, Mass: Addison-Wesley Pub. Co.

Alt, N. P., Goodale, B., Lick, D. J., & Johnson, K. L. (2019). Threat in the company of men: Ensemble perception and threat evaluations of groups varying in sex ratio. *Social Psychological and Personality Science*, *10*(2), 152–159. https://doi.org/10.1177/1948550617731498

Alt, N. P., & Phillips, L. T. (2022). Person perception, meet people perception: Exploring the social vision of groups. *Perspectives on Psychological Science*, *17*(3), 768–787. https://doi.org/10.1177/17456916211017858

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2019). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review. 23*(4), 307-331.

Balcetis, E., & Lassiter, D. G. (Ed.). (2010). *Social psychology of visual perception*. New York, NY: Psychology Press.

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

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674–679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull, & R. S. Wyer, Jr., (Eds.), A dual process model of impression formation, vol. 1: 1–36. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, *41*(3), 405–419. https://doi.org/10.1177/0146167214567153

Chwe, J. A. H., & Freeman, J. B. (2023). Trustworthiness of crowds is gleaned in half a second. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506231164703

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, P., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32,* 917–929. doi:10.1177/0146167206287721

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In D. K. Weisbert (Ed.), *Feminist legal theory: Foundations* (pp. 383-395). Philadelphia, PA: Temple University Press.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181–3192.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Esposito, E., & Calanchini, J. (2022). Examining selective migration as attitudinal fit versus Gay Migration. *Journal of Experimental Social Psychology*, *101*. https://doi.org/10.1016/j.jesp.2022.104307

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In Advances in Experimental Social Psychology (Vol. 23, pp. 1–74). Elsevier. https://doi.org/10.1016/S0065-2601(08)60317-2

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247–279. doi: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*(4), 673–690. https://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*(10), 1318–1331. https://doi.org/10.1177/0146167210378755

Friedman, H., & Zebrowitz, L. A. (1992). The contribution of typical sex differences in facial maturity to sex role stereotypes. *Personality and Social Psychology Bulletin*, *18*(4), 430–438. https://doi.org/10.1177/0146167292184006

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). "Ain't I a Woman?": Toward an intersectional approach to person perception and group-based harms. *Sex Roles, 59*, 392-403. https://psycnet.apa.org/doi/10.1007/s11199-008-9505-4

Goldenberg, A., Sweeny, T. D., Shpigel, E., & Gross, J. J. (2020). Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *Journal of Experimental Psychology: General*, *149*(3), 445–460. https://doi.org/10.1037/xge0000651

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

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Nature Current Biology, 17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Hehman, E., Sutherland, C. A., 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*(4), 513–529. https://doi.org/10.1037/pspa0000090

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1601785

Hess, U., Adams, R. B., Jr., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision, 9*(12), Article 19. https://doi.org/10.1167/9.12.19

Hester, N., & Hehman, E. (2023). Dress is a fundamental component of person perception. *Personality and Social Psychology Review*. Advance online publication. https://doi.org/10.1177/10888683231157961

Hester, N., Jones, B. C., & Hehman, E. (2020). Perceived femininity and masculinity contribute independently to facial impressions. *Journal of Experimental Psychology: General, 150*(6), 1147-1164.

Im, H. Y., Chong, S. C., Sun, J., Steiner, T. G., Albohn, D. N., Adams, R. B., & Kveraga, K. (2017). Cross-cultural and hemispheric laterality effects on the ensemble coding of emotion in facial crowds. *Culture and Brain*, *5*(2), 125–152. https://doi.org/10.1007/s40167-017-0054-y

Johnson, K. L., & Adams, R. B. (2013). Social Vision: An introduction. *Social Cognition*, *31*(6), 633–635. https://doi.org/10.1521/soco.2013.31.6.633a

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

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent research in social vision: an integrative approach to the determinants and consequences of social categorization. *Social and Personality Psychology Compass, 9*(1), 15-30. http://doi.org/10.1111/spc3.12147

Johnson, K. L., McKay, L. S., & Pollick, F. E. (2011). He throws like a girl (but only when he's sad): Emotion affects sex-decoding of biological motion displays. *Cognition, 119*(2), 265–280. https://doi.org/10.1016/j.cognition.2011.01.016

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly. *Psychological Science*, *16*(11), 890–897. https://doi.org/10.1111/j.1467-9280.2005.01633.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601-625.

Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology, 66,* 547–574. https://doi.org/10.1146/annurev-psych-010814-015025

Kemelmacher-Shlizerman, I., Suwajanakorn, S., & Seitz, S. M. (2014). Illumination-aware age progression. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* (pp. 3334–3341). IEEE Computer Society. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber1⁄46909822

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models, R package version 2.0-30. https://CRAN.R-project.org/package=lmerTest

Lamer, S. A., Sweeny, T. D., Dyer, M. L., & Weisbuch, M. (2018). Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *Journal of Experimental Psychology: General, 147*, 683-701. doi:10.1037/xge0000443

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Lei, R. F., Leshin, R. A., & Rhodes, M. (2020). The development of intersectional social prototypes. *Psychological Science, 31*(8), 911-926. https://doi.org/10.1177/0956797620920360

Lick, D. J., & Johnson, K. L. (2014). Perceptual underpinnings of antigay prejudice. *Personality and Social Psychology Bulletin*, *40*(9), 1178–1192. https://doi.org/10.1177/0146167214538288

Lick, D. J., Johnson, K. L., Rule, N. O., & Stroessner, S. J. (2019). Perceivers infer base rates from social context to judge perceptually ambiguous social identities. *Social Cognition, 37*(6), 596–623. https://doi.org/10.1521/soco.2019.37.6.596

Ma, D., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135. doi:10.3758/s13428-014-0532-5

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*(1), 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*(4), 383–401. https://doi.org/10.1207/s15327957pspr0804_4

Morgenroth, T., & Ryan, M. K. (2020). The effects of gender trouble: An integrative theoretical framework of the perpetuation and disruption of the gender/sex binary. *Perspectives on Psychological Science*, *16*(6), 1113–1142. https://doi.org/10.1177/1745691620902442

Oh, D., Buck, E. A., & Todorov, A. (2018). Revealing hidden gender biases in competence impressions of faces. *Psychological Science*, *30*(1), 65–79. https://doi.org/10.1177/0956797618813092

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oswald, F., & Adams, R. B. (2022). Feminist social vision: Seeing through the lens of marginalized perceivers. *Personality and Social Psychology Review*. https://doi.org/10.1177/10888683221126582

Oswald, F., Griffin, J. W., Weisbuch, M., & Adams, R. B. (2023). People watching: Social perception and the ensemble coding of bodies. *Journal of Nonverbal Behavior, 47,* 545-468.

Parkinson, A., Connell, H., Weeks, M., & Haberman, J. (2023, February 23-25). *Demonstrating ensemble perception of social status* [Poster presentation]. SPSP Annual Convention, Atlanta, GA. https://jasonmarchaberman.files.wordpress.com/2023/06/spsp-conference-poster_final_ap_hc.pdf.

Petsko, C. D., Rosette, A. S., & Bodenhausen, G. V. (2022). Through the looking glass: A lens-based account of intersectional stereotyping. *Journal of Personality and Social Psychology*, *123*, 763-787.

Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology, 114*(5), 766–785. https://doi.org/10.1037/pspi0000120

Phillips, L.T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior, 34*, 101-127.

Remedios, J. D., & Sanchez, D. T. (2018). Intersectional and dynamic social categories in social cognition. *Social Cognition: Special Issue on Intersectional and Dynamic Social Categories, 36,* 453-460. doi:10.1521/soco.2018.36.5.453

Schug, J., Alt, N. P., & Klauer, K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology, 56,* 121-125. https://doi.org/10.1016/j.jesp.2014.09.012

Sczesny, S., Spreemann, S., & Stahlberg, D. (2006). Masculine = competent? Physical appearance and sex as sources of gender-stereotypic attributions. *Swiss Journal of Psychology*, *65*(1), 15–23. https://doi.org/10.1024/1421-0185.65.1.15

Sternberg, B., Badea, C., & Rubin, M. (2023). Intersectional person perception: A scoping review of studies investigating the roles of category relationships and cognitive processes. *Social Cognition, 41*(2), 165-192.

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition, 14,* 247-276.

Sun, Z., He, Z., Zhang, G., Li, X., & Yu, W. (2020). Incidental learning of group trust: Predictive gaze cue matters. *Scientific Reports, 10*, 7789. Doi:10.1038/s41598-020-64719-5

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool for investigating social perception. *Social Psychological and Personality Science, 8*(5), 538-551.

Ward, C., Flowe, H., & Humphries, J. (2012). The Effects of Masculinity and Suspect Gender on Perceptions of Guilt. *Applied Cognitive Psychology, 26*(3), 482–488. doi:10.1002/acp.2823

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Yang, X., & Dunham, Y. (2019). Hard to disrupt: Categorization and enumeration by gender and race from mixed displays. *Journal of Experimental Social Psychology, 85*, 103893. https://doi.org/10.1016/j.jesp.2019.103893.

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8*(2), 161–185. https://doi.org/10.1521/soco.1990.8.2.161

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology, 98*(2), 175–189. https://doi.org/10.1037/a0017990

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology*, *28*(6), 1143–1152. https://doi.org/10.1037/0012-1649.28.6.1143

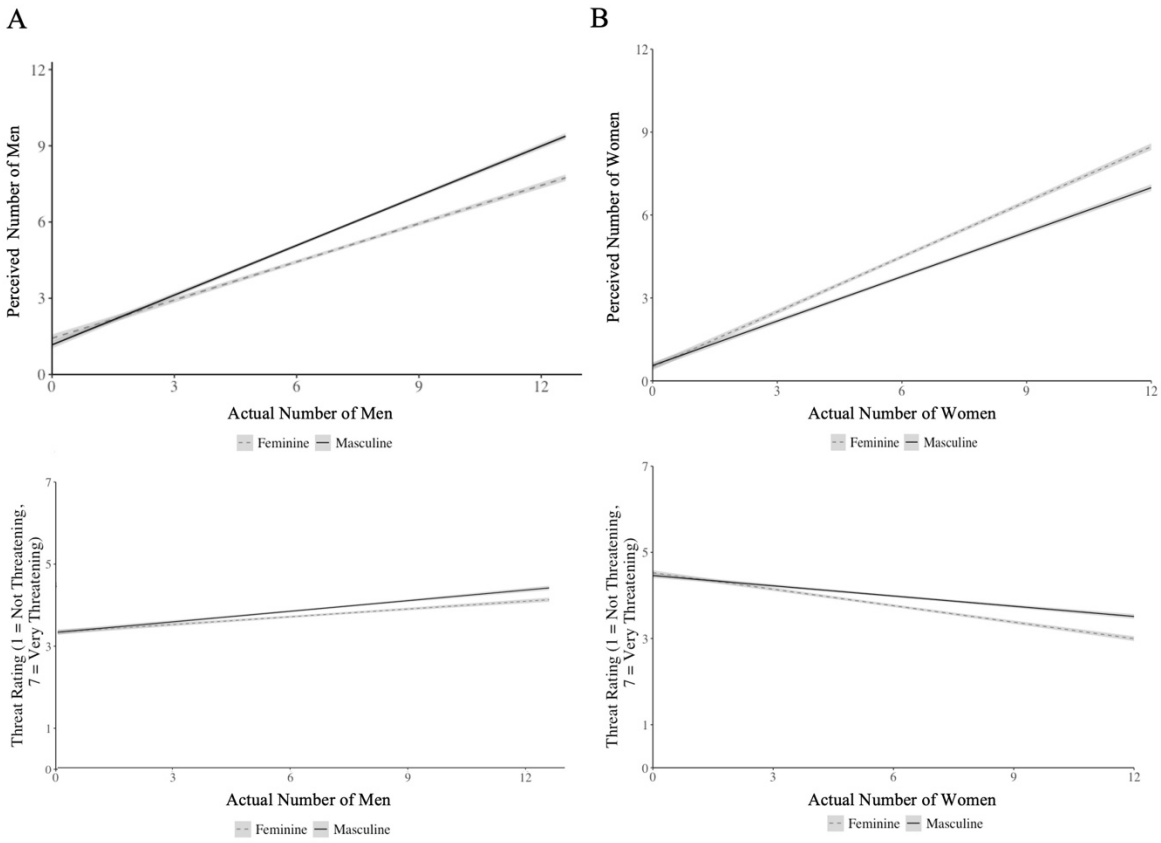

Figure 1. Results for Study 1 (Panel A) and Study 2 (Panel B) for Perceived Number of Men/Women and Perceived Threat. Error bars represent 95% confidence intervals.