Running head: INTERSECTION OF SEX AND EMOTION

Stereotypes bias visual prototypes for sex and emotion categories

Jeffrey A. Brooks, Ryan M. Stolier, & Jonathan B. Freeman

New York University

Word count: 5,856
Abstract word count: 145

Address correspondence to:

Jeffrey A. Brooks or Jonathan B. Freeman
Department of Psychology
New York University
6 Washington Place
New York, NY 10003
E-mail: jab1148@nyu.edu or jon.freeman@nyu.edu

Abstract

Recent models suggest that social categories are not perceived independently, but that they can facilitate or bias each other's perception due to incidentally shared stereotypes. To address the role of visual prediction in driving these intersectional effects in the domains of sex and emotion perception, three studies were conducted. Participants categorized visually obscured faces by sex and emotion, from which we produced reverse correlated prototype faces for each social category. These prototype faces were found to exhibit systematic biases in their visual appearance (Male-Angry, Female-Happy), as judged by independent raters. Moreover, this biased appearance in sex and emotion prototypes was related to the extent of a participant's stereotypical associations linking men to anger and women to happiness. Together, the findings suggest that stereotypes can bind social categories together at the level of visual prediction and offer new insights into current theoretical models of social perception.

Stereotypes bias visual prototypes for sex and emotion categories

Perceivers are able to effortlessly draw a remarkable amount of information about someone from even a brief glimpse of their face. Even a single-shot glance at an individual passing by on the street grants information about the person's sex, race, age, emotional state, as well as higher level social information like their potential personality characteristics. Moreover, perceivers display remarkable consensus in their categorizations of a target individual, even for such brief displays (Kenny, 1991). The seemingly automatic, obligatory nature of social categorization led to a tendency in the literature to discuss social categorization as a purely bottom-up perceptual process in which facial cues are read off the face, giving rise to automatic and stable categorizations that are independent of one another. However, recent work in social perception has challenged the idea of independence, showing that perception of an individual's social category memberships along one dimension (e.g. sex) can provide context for the perception of that individual's other social category memberships (e.g. emotion), facilitating or biasing social categorizations along other dimensions. The implication of this work is that social categories from seemingly unrelated dimensions might actually be inherently interdependent at the level of conceptual structure and thereby visual perception (Freeman & Ambady, 2011; Freeman & Johnson, 2016; Johnson & Freeman, 2010).

Social perception research has been increasingly informed by findings indicating that initial perceptions are subject to a variety of contextual influences, as well as "top-down" influences from pre-existing perceptual heuristics in the observer (Balcetis & Lassiter, 2010; Adams et al., 2010; Freeman & Johnson, 2016). Translating these observations into social perception has led to an appreciation of how aspects of the perceiver (including prior knowledge, conceptual knowledge, and stereotypes) can influence and in some cases systematically bias social perception. Stereotypes comprise a particularly consequential influence on social categorization, enabling seemingly independent social categories (like sex and emotion) to influence perceptions of one another during face processing when they share associated conceptual/stereotype content (Freeman & Johnson, 2016).

Stereotypes have long been assumed to guide social interaction, but research has accumulated to suggest that stereotypes also guide much earlier and more basic perceptual processes, before social categorization is complete. For example, studies using computer mouse-tracking are able to index participants' motor trajectories en route to clicking a response option in forced-choice categorization tasks (e.g. Dale, Kehoe, & Spivey, 2007; Freeman et al., 2008), and they are able to show how stereotypes can impact the temporal dynamics of social perception before categorizations are complete. In particular, deviation of a mouse trajectory toward a competing response provides an indirect measure of how much that category was activated during the perceptual process. The patterns of mouse deviation in social categorization tasks implicate the early role of stereotypes in social categorization, such that female-related stereotypes are partially activated while perceiving a slightly feminine male face (Freeman & Ambady, 2009) and that Black female faces partially activate the Male sex category because of the shared stereotypes between the Black and Male categories (Johnson, Freeman, & Pauker, 2012). Thus, social categories that incidentally share stereotypes can provide context for the perception of one another. These "intersectional" effects have been observed across several pairwise combinations of ostensibly unrelated social categories, such as sex and race (Black-Male and Asian-Female; Johnson, Freeman, & Pauker, 2012) and race and emotion (Black-Angry; Hugenberg & Bodenhausen, 2003; 2004).

An interesting case of intersectional effects exists in the domain of sex and emotion categories. Since men are stereotyped as more aggressive and women are stereotyped as more docile, intersectional effects have emerged such that male faces are biased to be perceived as angry and female faces are biased to be perceived as happy (Hess, Adams, & Kleck, 2004; Hess et al., 2000), leading to facilitated perceptions of angry males and happy females. These perceptual effects seem to follow stereotypic associations, but some research has also suggested that the perceptual interdependence of sex and emotion categories is due to confounded bottom-up cues, such that sexual dimorphism of the face causes male faces to incidentally contain visual cues to the emotion category “anger” and female faces to contain cues to the emotion category Happy (Becker et al., 2007).[1]

However, recent neuroimaging work has shown that neural response patterns associated with sex and emotion categories in the right fusiform gyrus (FG), a region centrally involved in face perception (Haxby , Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997), and orbitofrontal cortex (OFC), a region associated with top-down visual predictions and stereotype access (Bar et al., 2006; Freeman et al., 2010; Milne & Grafman, 2001), are systematically more similar when those categories share overlapping stereotype content, even after controlling for the visual similarity of faces from the categories in question (Stolier & Freeman, 2016). These findings suggest that stereotypes may play a general role in facilitating the intersection of social categories above and beyond any potential bottom-up overlap in those categories' associated facial cues.

Moreover, the fact that this similarity structure (i.e., Male more similar to Angry, and Female more similar to Happy) was observed in the OFC offers a hint about the specific cognitive role that stereotypes play in facilitating intersectional effects. The object recognition literature suggests that, in categorization and perceptual decision making tasks, the OFC is specifically recruited when incoming visual sensory information matches a pre-existing representation in memory, or a task-based expectation, like expecting to see face or house stimuli in an experimental block (Summerfield & Egner, 2009). Coarse early representations of visual input are sufficient to drive this predictive activity in the OFC, suggesting that the OFC is involved in exerting visual predictions about category membership before those categorizations have fully crystallized (Bar, 2003; Bar et al., 2006). The fact that similar patterns of results emerge in neuroimaging work on social categorization suggests that stereotypes may be involved in a similar predictive capacity during social perception. For example, early processing of cues to the sex category Female may activate visual predictions that are guided by stereotypes and their associated conceptual and behavioral heuristics. Since this predictive activity could take on some of the visual processing load, it may then be able to feed back and bias ongoing visual processing, facilitating the perception of other social categories that may not even be valid for a given face (Freeman & Johnson, 2016).

One behavioral method that has demonstrated the guiding role of expectation and prediction in social perception is reverse correlation (Dotsch & Todorov, 2012), in which stimuli are developed by partially obscuring a neutral base face with random patterns of visual noise. On a given trial, participants are tasked with deciding which of two faces belong to a given social category, even though the two faces in question are the identical base face, overlaid with different patterns of visual noise. Across several trials, researchers are able to statistically average the faces that each participant placed into a given social category, generating an idiosyncratic prototype for that category for that participant. The method yields a facial map of the facial features that participants expect to see for a given social category, making this method potentially productive for assessing predictive aspects of face perception (and social categorization in particular). These prototype faces are then combined into “grand average” faces for each category and used as stimuli in an independent rating task. Research using this method has found that the resulting prototypes are systematically warped by participants’ individual top-down predictions, such that highly prejudiced individuals yield more stereotypical and untrustworthy faces when they are asked to categorize by race (Dotsch et al., 2008), and individuals who have undergone a “minimal groups” manipulation yield more trustworthy reverse correlated faces when they are asked to categorize their in-group (Ratner et al., 2014). Thus, the technique produces faces that can be interpreted as an individual’s top-down prediction for a given social category or class of faces.

In the present research, we used the reverse correlation technique to assess the degree to which sex and emotion categories are related at the level of top-down visual predictions. If certain categories are perceptually entangled (e.g. the sex category Male and the emotion category Angry) then the visual prototypes produced through the reverse correlation technique for one category of faces (e.g. Male) may incidentally contain cues to the unrelated category (e.g. Angry) even though participants were not attending to the unrelated category at all. We addressed this question for sex and emotion categories by generating reverse-correlated faces for the sex categories Male and Female and emotion categories Angry and Happy. We subsequently used these faces as stimuli in independent sex and emotion rating tasks. In line with prior work on the intersection of social categories (Hess, Adams, & Kleck, 2004; Stolier & Freeman, 2016) we predicted that the grand average sex-category faces would be biased toward certain emotion categories and vice-versa (Male-Angry, Female-Happy). The reverse correlation technique is particularly advantageous since the visual noise used to obscure the base faces is randomly generated on each trial, making the final prototype face entirely data-driven. There is no opportunity for a priori predictions by the experimenters about which cues are most diagnostic of each category to impact the results, so that any category overlap that emerges through the task does so in an unconstrained manner. Across 3 studies, we tested whether visual prototypes of sex and emotion categories intersect in a stereotypical fashion. Study 1 provided preliminary evidence for this possibility, which Studies 2 and 3 replicated with larger samples and an additional base face stimulus and, most importantly, specifically linked to inter-individual variability in stereotypical associations.

Study 1

In this preliminary study, we used the reverse correlation technique to test the possibility of a systematic bias in the visual prototypes of sex and emotion categories consistent with stereotypes (Male-Angry and Female-Happy).

Method

Participants

A total of 36 participants completed the study. Unfortunately, an *a priori* power analysis was not conducted for this preliminary study, but note that Studies 2 and 3 replicate the present study's pattern of results with larger samples. Half of the participants (0 female) completed the sex reverse correlation task (M_{age} = 33.29, 55.55% White, 27.78% Black, 16.67% other) and the other half (0 female) completed the emotion reverse correlation task (M_{age} = 36.11, 88.89% White, 11.11% other). Thirty independent raters (17 female) completed the subsequent sex rating task (M_{age} = 41.73, 83.33% White, 13.33% Black, 3.33% Asian) and 30 independent raters (16 female) completed the subsequent emotion rating task (M_{age} = 35, 80% White, 6.667% Black, 6.667% Asian, 6.667% other). No participants or independent raters were excluded and our sample size was determined by the sample sizes reported in other studies using the reverse correlation technique and assessing the relationship between sex and emotion categories (Dotsch et al., 2008; Hess, Adams, & Kleck, 2004). Participants and independent raters completed the tasks through Amazon's Mechanical Turk in exchange for monetary compensation.

Procedure

Following the reverse correlation procedure implemented in other research (Dotsch et al., 2008; Dotsch & Todorov, 2012), participants were presented with two side-by-side faces on each trial. These faces were the same base face, with different patterns of randomly generated visual noise placed over each. The sex classification task was divided into 200 trials that required participants to choose the face that looked more Male and 200 trials that required participants to choose the face that looked more Female. The emotion classification task was divided into 200 trials that required participants to "Choose the face that looks more Angry" and 200 trials that required participants to

“Choose the face that looks more Happy”. Following the data preprocessing approach taken in other reverse correlation research (Dotsch & Todorov, 2012), for each task we averaged the face selected on each trial by each participant, resulting in classification images (CIs) for the categories Anger and Happiness (for the 18 participants who did the emotion reverse correlation task) and the categories Male and Female (for the 18 participants who did the sex reverse correlation task), resulting in a total of 72 CIs (Figure 1).

In the subsequent rating task, independent groups of participants rated each of the 72 CIs along sex or emotion categories. For sex ratings, participants were instructed to rate how Male or Female the face looked using a sliding scale from 1 = Male to 7 = Female. For emotion ratings, participants were instructed to rate how Angry or Happy the face looked using a sliding scale from 1 = Happy to 7 = Angry.

In all studies, we report all measures and conditions collected in the task, as well as all participant exclusions.

Stimuli

The base face for the reverse correlation task was created by averaging 384 face stimuli (from Stolier & Freeman, 2016) that varied in sex (Female, Male), race (Asian, Black, White), and emotion categories (Angry, Happy). The faces were created using FaceGen modeler, a 3D face modeling software that allows for precise manipulation of social category cues. Thirty-two faces from each possible combination of sex, race, and emotion categories (e.g. Angry-Asian-Male, Happy-Black-Female) were included in the average.

Results and Discussion

Consistent with our predictions, the Male CIs were reliably rated as Angrier (*M* = 3.982, *SD* = 1.453) and the Female CIs were rated as Happier (*M* = 3.265, *SD* = 1.265), *t*(29) = 7.05, *p* < .0001, *d* = 1.287. Likewise, the Angry CIs were rated as more male (*M* = 3.252, *SD* = 1.748) and the Happy CIs were rated as more female (*M* = 4.043, *SD* = 1.840), *t*(29) = 6.095, *p* < .0001, *d* = 1.112 (Figure 3). Together, these results suggest that sex and emotion cues provide context for the perception of one another in a manner potentially consistent with stereotypes. Next, we aimed to replicate these findings with a larger sample, and importantly, provide evidence specifically linking these intersectional effects of sex and emotion to relevant stereotypical associations.

Study 2

In Study 1 we found that reverse-correlated images of sex categories Male and Female and the emotion categories Angry and Happy were biased to be perceived along presumably orthogonal dimensions in a manner consistent with stereotypes (Male-Angry and Female-Happy). However, it is difficult to disentangle the relative contributions of bottom-up factors (e.g. overlap in facial cues related to the categories Male and Angry) vs. top-down factors (e.g. overlap in conceptual/stereotype content for these categories) in these results. In Study 2, we repeated the design of Study 1 in a larger sample for which we also collected stereotype content data for the categories Male, Female, Angry, and Happy. This measure allows us to determine how much conceptual/stereotype overlap there is between categories for each participant in the initial reverse correlation task. In turn, this allows us to determine the degree to which the results in the subsequent rating task were due to top-down conceptual/stereotype overlap.

As in Study 1, we predicted that CIs for sex and emotion categories would be biased to be perceived along orthogonal dimensions in a manner consistent with stereotypes (Male-Angry and Female-Happy). We also predicted that the strength of this effect would be related to the extent to which a participant harbored stronger stereotypical associations between sex and emotion categories. For example, we predicted that someone with high conceptual/stereotype overlap between the categories Male and Angry would produce a Male (or Angry) CI that was angrier (or more masculine) in appearance than someone with less overlap.

Method

Participants

Because this study focused on individual differences and sought to replicate Study 1's findings with a larger sample, we aimed to recruit a total of 150 participants. A priori power analysis was not conducted as we were not aware of any previously reported effect size for the tests of individual differences. A total of 152 participants completed the study. Seventy-five participants (41 female) completed the initial emotion reverse correlation task (M_{age} = 40.4). Four participants were excluded for failing to follow instructions on the stereotype content portion of the task, resulting in a sample of 71. Seventy-seven participants (52 female) completed the initial sex reverse correlation task (M_{age} = 40.62). No participants were excluded from this task.[2] Twenty-nine independent raters (17 female) completed the subsequent emotion rating task (M_{age} = 36.76, 75.86% White, 6.9% Black, 17.24% Other) and 28 independent raters (15 female) completed the subsequent sex rating task (M_{age} = 35.38, 65.38% White, 19.23% Black, 7.69% Asian, 7.69% Other). One rater was excluded from the sex rating task for not following instructions, resulting in a set of 27 raters.

Procedure

______In the initial reverse correlation task, participants classified visually obscured faces by sex or emotion categories in a procedure identical to that in Study 1. The data preprocessing approach was also identical, resulting in CIs for the categories Anger and Happiness (for the 71 participants who did the emotion reverse correlation task) and the categories Male and Female (for the 77 participants who did the sex reverse correlation task), resulting in a total of 296 CIs. Participants in this phase of the study also completed a conceptual/stereotype content task, where they rated a battery of 96 traits (previously used in Stolier & Freeman, 2016) on how much they theoretically applied to an individual from each of the four sex/emotion categories (e.g., hostile, timid). See Footnote 1. Specifically, participants were asked to judge how well each word or phrase stereotypically describes each category (Male, Female, Angry, Happy) on a scale from 1 ("non-stereotypic") to 7 ("very stereotypic"). By correlating each participant's response vectors for each category (i.e., the 96 Likert ratings), we were able to estimate each participant's own conceptual/stereotype overlap for each pair of categories (e.g. Male and Angry, Female and Happy). See Footnote 1.

In the subsequent rating task, independent groups of raters rated each of the 296 CIs along either sex or emotion categories in a procedure identical to that in Study 1.

Stimuli

The base face for the reverse correlation task was identical to that used in Study 1.

Results and Discussion

Replicating Study 1's effects, Male CIs were perceived as Angrier than Happy (*M* = 4.09, *SD* = 0.31) and Female CIs were perceived as Happier than Angry (*M* = 3.77, *SD* = 0.32), *t*(28) = 3.99, *p* < .001, *d* = 1.01. Conversely, Angry CIs were rated as more Male than Female (*M* = 3.48, *SD* = 0.91) and Happy CIs were rated as more Female than Male (*M* = 3.78, *SD* = 0.84), *t*(26) = 3.74, *p* < .001, *d* = 0.3

To assess the contribution of participants' stereotypical associations in their sex and emotion CIs, we conducted a regression analysis testing whether overlapping conceptual/stereotype knowledge (e.g., 'hostile' shared by both Male and Angry categories) reliably predicted the extent of bias in a participant's CIs (e.g., for an emotion-task participant: how masculine their Angry CI and feminine their Happy CI appeared). For each participant across both reverse correlation tasks, we regressed the independent ratings of their two CIs (average male-female rating for emotion CIs, and average angry-happy rating for sex CIs) onto the relevant hypothesized conceptual/stereotype overlap, as detailed below.

For each participant and for each of their two CIs, we computed a score representing the relevant conceptual/stereotype overlap we hypothesized given the particular CI category, specifically: stereotypically-congruent overlap/similarity adjusted for stereotypically-incongruent overlap/similarity. For example, for a participant who completed the emotion task, for their Angry CI the stereotype score was computed as Angry-Male overlap/similarity (i.e., Pearson correlation), minus Angry-Female overlap/similarity (i.e., Pearson correlation). For their Happy CI, the score was computed as the Happy-Female correlation minus the Happy-Male correlation. For a participant who completed the sex task, for the Male CI it was computed as the Male-Angry correlation minus the Male-Happy correlation, and for the Female CI computed as the Female-Happy correlation minus the Female-Angry correlation. Thus, when conceptual knowledge of a pair of two categories (e.g., Angry-Male) was rated highly similarly on the 96 traits, that pair's correlation would approach 1, and when rated highly dissimilarly the correlation would approach -1, consistent with previous work using this kind of conceptual/stereotype content task (Stolier & Freeman, 2016). Therefore, the stereotype score for each participant and for each of their two CIs could range from 2 (very strong Angry-Male/Happy-Female bias), to 0 (no bias), to -2 (very strong counter-hypothetical Angry-Female/Happy-Male bias). This score served as the main independent variable in this analysis.

Independent ratings along the orthogonal sex/emotion dimension served as the dependent measure for each participant's two CIs. Sex (male-female) and emotion (angry-happy) ratings were centered and re-coded such that 3 indicates stereotypical congruency and -3 indicates stereotypical incongruency in visual appearance. Thus, positive values indicate a Male CI appears angry, a Female CI appears happy, an Angry CI appears masculine, and a Happy CI appears feminine; negative values, on the other hand, indicate the opposite (counter-hypothetical bias). Thus, coded in this fashion, if a participant's conceptual/stereotype knowledge linking sex and emotion categories is related to the biased appearance of their CIs, we should expect a strong positive relationship between the stereotype overlap scores and the stereotypical congruency of their CIs.

Due to the nested structure of the data (CIs nested within participant), we used a multi-level regression approach using generalized estimating equations (GEE) (Zeger &

Liang, 1986). Task was included as an independent variable (-0.5 = emotion, 0.5 = sex), as well as the interaction between task and stereotype overlap, to assess the possibility of stronger stereotype-driven biases for sex vs. emotion tasks. Consistent with our hypothesis, there was a significant main effect of stereotype overlap on the level of bias in a participant's CIs, such that individuals with a higher degree of overlap between the categories in question (Male-Angry and Female-Happy) produced CIs exhibiting a corresponding degree of bias in visual appearance, $B = 0.23$, $SE = 0.06$, 95% CI [0.11, 0.35], Wald $Z = 3.79$, $p < .001$. Interestingly, there was also a significant main effect of task, such that participants who completed the sex reverse correlation task were more likely to produce CIs that were biased along the orthogonal (emotion) dimension than participants who completed the emotion task, $B = 0.459$, $SE = 0.06$, 95% CI [0.34, 0.58], Wald $Z = 7.64$, $p < .0001$. However, there was no interaction between task and stereotype overlap, indicating that the influence of stereotypes did not change as a function of which task the CIs were generated from, $B = -0.052$, $SE = 0.12$, 95% CI [-0.29, 0.18), Wald $Z = -0.43$, $p = .67$.

Thus, in Study 2 we replicated the results of Study 1 and found that reverse-correlated prototypes for sex and emotion categories were perceived by independent raters in a manner consistent with stereotypes. Further, we found that CIs were more likely to be perceived in a stereotype-congruent manner if the participant from whom we produced the prototype image harbored a high degree of conceptual/stereotype overlap between sex and emotion categories. This pattern of results suggest that emotion and sex categories provide context for the perception of one another, but importantly, also that this effect is driven by the degree of overlap in stereotype knowledge between the categories in question.

Study 3

Here, we repeated the design of Study 2, but used a different base face for the initial reverse correlation task. Since the base face we used in the previous task was averaged from stimuli used in a separate study (and containing a blend of cues from multiple race, sex, and emotion categories), it is difficult to know precisely how unconstrained the initial categorizations were in the reverse correlation task. Here we used a completely neutral base face used in prior research (Dotsch et al., 2008), allowing us to demonstrate the generality of our results while also ensuring that the initial reverse correlation task was not constrained by any incidental aspects of the stimuli used.

Method

Participants

A total of 159 participants completed the study. Seventy-seven participants (50 female) completed the initial emotion reverse correlation task (M_{age} = 39.97). Three participants were excluded for failing to follow instructions during the stereotype content portion of the task, resulting in a sample of 74. Eighty-two participants (56 female) completed the initial sex reverse correlation task (M_{age} = 39.99). Six participants were excluded for failing to follow instructions during the stereotype content portion of the task, resulting in a sample of 76. See Footnote 2. Thirty-two independent raters (21 female) completed the subsequent emotion rating task (M_{age} = 40.31, 68.75% White, 15.63% Black, 9.34% Asian, 6.25% Other) and 29 independent raters (19 female) completed the subsequent sex rating task (M_{age} = 40.76, 79.31% White, 6.9% Black, 6.9% Asian, 6.9% Other). No raters were excluded from these tasks.

Procedure

In the initial reverse correlation task, participants classified visually obscured faces by sex or emotion categories in a procedure identical to that in Studies 1 and 2. Participants in this phase of the task also completed a task designed to measure the conceptual/stereotype content of the categories Male, Female, Happy, and Angry. Collection and analysis of this data was performed with the same procedure as Study 2.

In the subsequent rating task, an independent group of raters rated each of the 300 CIs along sex and emotion categories in a procedure identical to that performed in Studies 1 and 2.

Stimuli

The base face used in this study was the same image used in the first published study to use the reverse correlation technique (Dotsch et al., 2008). As described by Dotsch et al. (2008), the image is the neutral male mean from the Averaged Karolinska Directed Emotional Faces database (Lundqvist & Litton, 1998).

Results

Replicating Studies 1 and 2, Male CIs were reliably rated as Angrier than Happy (M = 4.82, SD = 0.65) and the Female CIs were rated as Happier than Angry (M = 4.15, SD = 0.57), $t(31)$ = 6.57, p < .001, d = 1.09. Similarly, the Angry CIs were rated as more Male than Female (M = 2.78, SD = 0.90) and the Happy CIs were rated as more Female than Male (M = 4.36, SD = 0.87), $t(28)$ = 12.21, p < .001, d = 1.78.

We conducted an analogous regression analysis as used in Study 2 to predict biases in the generated CIs from participant's stereotype overlaps. Once again, we observed a significant main effect of stereotype overlap, *B* = 0.71, *SE* = 0.092, 95% CI [0.53, 0.89], Wald *Z* = 7.71, *p* < .0001. There was also a significant main effect of task as in Study 2, indicating greater bias in the appearance of sex-task CI than emotion-task CIs, *B* = 1.079, *SE* = 0.069, 95% CI [0.94,1.21], Wald *Z* = 15.62, *p* < .0001. Unlike Study 2, however, we also observed a significant interaction between task and stereotype overlap, *B* = -0.88, *SE* = 0.18, 95% CI [-1.24, -0.52], Wald *Z* = -4.79, *p* < .0001. Simple slope analyses indicated that, while the positive relationship between stereotype overlap and biased appearance of the CIs was strong and robust across both tasks, it was somewhat stronger for emotion-task CIs (simple *B* = 0.27, *SE* = 0.062, 95% CI [0.15, 0.39], Wald *Z* = 4.31, *p* < .0001) than sex-task CIs (simple *B* = 1.15, *SE* = 0.17, 95% CI [0.81, 1.49], Wald *Z* = 6.65, *p* < .0001) tasks.

Thus, we replicated the results of the previous studies and also established a greater generality, showing that conceptual/stereotype overlap influences face categorization at the level of top-down visual prediction and that this relationship holds across multiple stimulus sets.

General Discussion

Behavioral and neuroimaging evidence has now repeatedly shown that seemingly unrelated social categories can become interconnected when they share similar stereotypes, but an understanding of the exact nature of this relationship has remained elusive. The current work used reverse correlation to determine whether sex and emotion categories bound by stereotype knowledge are indeed interconnected at the level of top-down visual expectation and prediction, since reverse correlated faces indicate the visual cues that perceivers expect to see for a given social category. Consistent with our predictions, the reverse-correlated “grand average” faces for the sex categories Male and Female contained cues to particular emotion categories (Male-Angry, Female-Happy), in a manner consistent with stereotypic expectations. Moreover, we also showed that the stereotype overlap of participants in the reverse correlation task predicted the degree to which their CIs were biased in their perception along orthogonal dimensions, suggesting that conceptual (i.e. stereotype) overlap does indeed strongly contribute to perceptual overlap between sex and emotion categories. Our findings demonstrate that social categories are able to become entangled at the level of top-down visual predictions, such that visual “prototypes” for certain social categories contain visual expectations about the other categories that a face is likely to belong to. These visual predictions are used to rapidly make sense of visual input that is oftentimes impoverished, lacking, or brief, enabling them to have a broad impact on momentary perceptions as well as behavior.

Importantly, this overlap between category dimensions occurred completely spontaneously in an unconstrained task. Prior work has only been able to address the role of shared stereotypes in category intersection through correlational techniques, for example measuring perceptual and conceptual (i.e. stereotype) similarity between social categories separately (Hess, Adams, & Kleck, 2004; Stolier & Freeman, 2016), or by having participants specifically reflect on how male or female the prototypical angry or happy face looks (Becker et al., 2007). The present work shows that these relationships spontaneously emerge on their own in a pattern consistent with prior work on stereotype overlap. This is particularly striking since the first task only required participants to attend to a given category (sex or emotion) in isolation. Our findings were also bidirectional: the grand average sex faces were each biased toward a particular emotion category and the grand average emotion faces were each biased toward a particular sex category (Male-Angry, Female-Happy), suggesting a relationship between sex and emotion categories such that they serve as context for each other's perception.

In Studies 2 and 3, we also showed that the category overlap effects observed in Study 1 were largely predicted by individual differences in stereotype overlap. For example, an individual who harbored a high degree of conceptual (stereotype) similarity between the categories Male and Angry was more likely to produce a Male CI that would be perceived as Angry. Prior work on category overlap has shown that perceptual overlap between sex and emotion plays out in a manner consistent with what would be expected from stereotypes, but in the present work we show that stereotypes are indeed a strong predictor of these overlap effects. The fact that these effects emerged in a reverse-correlation task indicates that stereotypes are able to influence visual predictions and expectations about a face's category membership, consistent with prior theoretical models (Freeman & Johnson, 2016; Freeman & Ambady, 2011).

We should note that, while the reverse correlation technique offers numerous advantages in understanding visual predictions, it is less clear what underlying representations those predictions might reflect. Prior work has commonly interpreted the reverse-correlated face stimuli as producing a given participants' "mental representation" or "visual representation" of the categories under investigation, but in the context of the present work such an interpretation would be overstated. On each trial, participants are tasked with choosing which of two visually obscured faces look most like the category in question, potentially leading to a visual search process for features that participants *expect* to see for a given face. Across many trials and participants, this produces facial maps of the cues participants expect to see. In our case, this makes the method particularly useful for studying visual predictions, but how these predictions relate to the way facial prototypes are stored in memory is uninformed by the technique. Future work integrating the reverse correlation technique with eye-tracking or neuroimaging measures could better speak to the cognitive interpretation of the technique and could refine interpretations of existing results. Nevertheless, the present work further characterizes the level of processing at which sex and emotion category intersection occurs, showing that stereotypes can bias face processing at the level of visual prediction and expectation, regardless of task context or overlapping bottom-up input.

These results highlight the top-down nature of intersectional effects with sex and emotion categories, but certainly do not exclude the possibility of co-existing bottom-up influences (i.e., direct physical overlap; Becker et al., 2007). However, regardless of the mechanism underlying intersecting sex and emotion categories, prior research has shown that the ability of sex and emotion categories to bias the perception of one another can have consequential effects on behavior. For example, over-perceiving anger in men and happiness in women may reinforce or possibly exacerbate already-existing behavioral biases. Moreover, such biases may make expressions of anger in women particularly noteworthy and norm-violating, resulting in negative evaluative consequences (Hess et al., 2000; 2009; 2010). Thus, these results also have implications for recent insights that deeply engrained perceptual associations may contribute to evaluative bias (Xiao, Coppin, & Van Bavel, 2016; Freeman, Pauker, & Sanchez, 2016). Since our results were also driven by individual differences in the degree of stereotype overlap in perceivers, this suggests that stereotype-related interventions (e.g. stereotype learning) could potentially be a target for interventions in this sort of perceptual bias.

In summary, we provide evidence that the intersection of sex and emotion categories in perception is driven, at least in part, by top-down visual prediction and expectation, which is influenced by individual differences in stereotype overlap. In particular, we found that sex and emotion categories provide context for one another such that male faces are biased to be perceived as angry, angry faces are biased to be perceived as male, female faces are biased to be perceived as happy, and happy faces are biased to be perceived as female. Moreover, this pattern was driven by stereotypes, bolstering prior work suggesting that stereotypes have an early influence on perception. As research accumulates to suggest the role of early perceptual processes in facilitating and maintaining bias, further work could help translate these insights into potential interventions to reduce the effects of stereotyping.

Author Note

This work was funded in part by research grant NSF-BCS-1423708 (J.B.F).

Footnotes

Footnote 1. While facial emotion is a category based on facial cues that are often dynamic, emotion perception does exhibit categorical perception effects in a manner similar to other social category perceptions such as sex. For this reason, current perceptually-driven models of social categorization consider both sex and emotion as inherently similar social categories that each have associated facial features and associated conceptual knowledge (Freeman & Johnson, 2016; Freeman & Ambady, 2011). In addition, for our purposes here, we are using the term "stereotype" to refer to stereotypical associations between sex and emotion categories. Since we are defining stereotypes as conceptual knowledge associated with a social category, this includes emotion categories as well, consistent with the models cited above. We do not imply any theoretical distinction between stereotypes vs. conceptual knowledge in this work.

Footnote 2. Due to an unfortunate error in data collection, demographic data on self-reported race/ethnicity of the participants in the initial reverse correlation task of Studies 2 and 3 were not collected.

References

Adams, R. B., Ambady, N., Nakayama, K., & Shimojo, S. (2010). *The science of social vision*. New York, NY: Oxford University Press.

Balcetis, E., & Lassiter, D. (2010). *The Social Psychology of Visual Perception*. New York, NY: Psychology Press.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600-609.

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449-454.

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.

Dale, R., Kehoe, C., & Spivey, M.J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15-28.

Dotsch, R., Wigboldus, D.H.J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, 19(*10*), 978.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(*5*), 562-571.

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. *Advances in Experimental Social Psychology*, *23*, 1-74.

Freeman, J.B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*, 1183-1188.

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.

Freeman, J.R., & Johnson, K.L. (2016). More than meets the eye: split-second social perception. *Trends in Cognitive Sciences*, *20*(5), 362-374.

Freeman, J.B., Pauker, K., & Sanchez, D.T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*.

Freeman, J.B., Rule, N.O., Adams, R.B., Jr., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex*, *20*, 1314-1322.

Galinsky, A.D., Hall, E.V., & Cuddy, A.J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, *24*(4), 498-506.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. A. (2000). The distributed human neural system for face perception. Trends in Cognitive Sciences, 4, 223–233.

Hess, U., Adams, R.B., & Kleck, R.E. (2004). Facial appearance, gender, and emotion expression. *Emotion*, *4*, 378-388.

Hess, U., Adams, R.B., Grammer, K., & Kleck, R.E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision*, *9*, 1-8.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R.E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition & Emotion*, *14*, 5.

Hess, U., Thibault, P., Adams, R.B., & Kleck, R.E. (2010). The influence of gender, social roles and facial appearance on perceived emotionality. *European Journal of Social Psychology*, 40(*7*), 1310-1317.

Hugenberg, K., & Bodenhausen, G.V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640-643.

Hugenberg, K., & Bodenhausen, G.V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342-345.

Johnson, K.L., & Freeman, J.B. (2010). A "New Look" at person construal: Seeing beyond dominance and discreteness. In E. Balcetis & D. Lassiter (Eds.), *The Social Pscyhology of Visual Perception*. New York: Psychology Press.

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*, 116-131.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience, 17(11), 4302-4311.

Kenny, D.A. (1991). A general model of consensus and accuracy in interpersonal perception. Psychological Review, 98, 155-163.

Lundqvist, D., & Litton, J.E. (1998). The Averaged Karolinska Directed Emotional Faces—AKDEF [CD ROM]. Stockholm: Karolinska Institute.

Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. The Journal of Neuroscience, 21(12): RC150.

Ratner, K.G., Dotsch, R., Wigboldus, D., van Knippenberg, A., & Amodio, D.M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*, 897-911.

Stolier, R.M., & Freeman, J.B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(*9*), 403-409.

Xiao, Y.J., Coppin, G., & Van Bavel, J.J. (2016). Perceiving the world through group-colored glasses: A perceptual model of intergroup relations. *Psychological Inquiry*, *27*, 255-274.