Shifts in facial impression structures across group boundaries

Youngki Hong[1] & Jonathan B. Freeman[1]

[1]Columbia University

Author Note

Youngki Hong, Department of Psychology, Columbia University, New York, NY, USA; Jonathan B. Freeman, Department of Psychology, Columbia University, New York, NY, USA

Correspondence concerning this article should be addressed to Youngki Hong, Department of Psychology, Columbia University, Schermerhorn Hall, New York, NY 10027. E-mail: youngki.hong@columbia.edu

Abstract

Facial impressions have long been argued to be driven by two independent dimensions of trustworthiness and dominance. However, in an intergroup context, we reasoned that these dimensions may shift predictably and become more positively correlated for ingroup members, yet negatively correlated for outgroup members, due to dominance signaling outgroup threat and/or ingroup prosociality. In two studies, we examined how the two dimensions shift across minimal group boundaries for White targets. In Study 1, core dimensions of trustworthiness and dominance became intertwined with each other differently for ingroup and outgroup targets. In Study 2, stronger stereotypic beliefs that trustworthiness ≈ dominance for ingroup than outgroup mediated the shifts in facial impression dimensions. This work advances our understanding of facial impression and intergroup bias by showing that the facial impression dimensions are not fixed but may shift across group boundaries and that such shifts occur above and beyond simple ingroup favoritism.

Keywords: person perception, minimal group paradigm, faces, trustworthiness, dominance

Abstract word count: 149/150

Shifts in facial impression structures across group boundaries

People spontaneously infer a broad range of personality traits from faces (Klapper et al., 2016; Oosterhof & Todorov, 2008). Although such facial impressions are typically inaccurate (Jaeger et al, 2021; Rule et al., 2013), information inferred from faces strongly influences a wide range of social decision-making, including some with dire consequences such as criminal sentencing (Duarte et al., 2012; Hassin & Trope, 2000; Todorov et al., 2005; Wilson & Rule, 2015). Facial impressions tend to be highly consistent across multiple perceivers (Hehman et al., 2017; Todorov et al., 2008), social groups (Cogsdill et al., 2014), and world regions and cultures (Jones et al., 2021). Oosterhof and Todorov's (2008) influential model of facial impressions has argued that, of all the possible trait judgments we can make, facial impression structure can be boiled down to two independent dimensions of trustworthiness and dominance. They argue that these dimensions are universal and arose due to their functional importance to survival: trustworthiness indicates people's good or bad intentions, whereas dominance indicates their ability to enact those intentions.

Although facial impressions show consistency across perceivers, a growing body of research has provided evidence that diversity in perceiver and target characteristics, as well as top-down processes among perceivers, also impact perceivers' facial impression structures. Perceiver characteristics are estimated to contribute a significant proportion of the variance in facial impressions overall (Hehman et al., 2017, 2019; Xie, et al., 2019). Other research has demonstrated that the core dimensions underlying facial impression structures – trustworthiness and dominance – can shift or disappear entirely depending on various factors, such as targets' social group memberships (Collova et al., 2019; Oh et al., 2020; Sutherland et al., 2015), perceivers' idiosyncratic beliefs about traits (Stolier et al., 2018, 2020), racial and gender

stereotypes (Xie et al., 2021), and perceivers' cultural environment (Oh et al., 2022; Sutherland et al., 2020). These findings suggest that the idea of a fixed and universal architecture for facial impressions does not fully account for more malleable and dynamic facial impression structures that predictably vary across many characteristics of who is judging and who is being judged.

While recent research has demonstrated variations in facial impression structures by target race and gender, it is unclear how ingroup and outgroup status affects facial impression structures. Extant research on intergroup face processing has demonstrated a similar ingroup favoritism that is observed in other kinds of intergroup judgments (Tajfel et al., 1971). For example, people associate more favorable traits such as trustworthiness with ingroup faces more so than outgroup faces (Hong & Ratner, 2021; Hutchings et al., 2021; Ratner et al., 2014), perceive ingroup members' emotional expressions as more positive than the same emotional expressions displayed by outgroup members (Beaupré & Hess, 2003; Dunham, 2011; Lazerus et al., 2016), and allocate greater cognitive resources to processing ingroup versus outgroup faces (Hong et al., 2022; Hugenberg & Corneille, 2009) that results in better recognition memory of ingroup faces (Bernstein et al., 2007; Hugenberg et al., 2010).

An unresolved question, however, is whether group membership affects the structure of facial impressions above and beyond any well-documented evaluative bias due to ingroup favoritism. For instance, dominance on an outgroup member's face might be perceived as outgroup members' ability to enact "bad" intentions and thereby pose a threat to ingroup resources, culture, and values (Esses et al., 1993). This would result in perceived dominance being relatively more negatively related to perceived trustworthiness for outgroup members. Conversely, dominance on an ingroup member's face might be perceived as their ability to enact "good" intentions because they are expected to provide support and resources (Brewer, 2004),

and thus perceived dominance would be relatively more positively related to perceived trustworthiness for ingroup members. Given the bidirectional conceptual associations between traits that scaffold facial impressions structure (Stolier et al., 2018, 2020), if these varied between ingroup and outgroup, then the reverse may be true as well. Higher perceived trustworthiness would lead to the perception of relatively higher dominance for ingroup compared to outgroup members. The stereotype content model also argues that there are two core dimensions underlying stereotype content – warmth and competence – and that ingroup members are often stereotyped as both high warmth and high competence, whereas outgroup stereotypes vary across different groups (Cuddy et al., 2009; Fiske et al., 2002). It is conceivable that targets' ingroup/outgroup status would similarly shift overall facial impression structures.

In the current research, we examined the effects of targets' ingroup/outgroup status on overall facial impression structures. To control any effects of preexisting stereotypes and prejudices people might have about real-world groups, we used the minimal group paradigm (Tajfel et al., 1971). In Study 1, we examined whether overall facial impression structures differ across minimal ingroup and outgroup. While ingroup/outgroup motivational processes would likely to play a role, a more proximal mechanism facilitating such differences may be perceivers' different conceptual associations about the ingroup vs. outgroup. Thus, we examined whether shifts in conceptual knowledge structures across ingroup and outgroup could explain corresponding shifts in facial impression structures. Study 2 examined the differential relationships between trustworthiness and dominance across ingroup and outgroup by manipulating rather than measuring trait-related facial appearance. We also examined the intermediary role of conceptual knowledge in the shifts in facial impression structures across

ingroup and outgroup. All data, materials, and analysis scripts are available at https://osf.io/6xcvg/.

Study 1

In Study 1, we predicted that the overall structures of ingroup and outgroup facial impressions would diverge due to distinct conceptual knowledge structures for ingroup and outgroup, which involves the two core dimensions underlying facial impressions and conceptual knowledge, trustworthiness and dominance, shifting across group boundaries. We test this using representational similarity analysis (RSA), an approach that has been used to assess structural shifts in facial impressions (Stolier et al., 2018) including in minimal group contexts (Hong & Ratner, 2021). We then characterized such shifts by examining changes in the relationships between core dimensions of trustworthiness and dominance across ingroup and outgroup. Because previous research finds that trustworthiness and dominance are weakly negatively related (r=-.20; Oosterhof & Todorov, 2008)[1], we expected that these dimensions would be somewhat negatively related overall at baseline but significantly less so for ingroup and more so for outgroup.

Method

Participants. Our target sample size was 265 participants based on the sample size necessary to detect a small effect size (Cohen's d = .20) at 90% power using a paired-samples t test. We rounded up our target size and recruited 300 White participants from Prolific to participate in an online study about how people make social judgments. We recruited White participants only to control for any effects of preexisting racial stereotypes and prejudices. Participants who did not have any variability in their responses or failed attention checks were

[1] The CFD norming data of the faces used in this study showed a somewhat stronger negative relationship between trustworthiness and dominance (r=-.49; Supplemental Analysis, S1).

excluded from the analyses. After exclusion, our final sample size was 257 participants (M_{age} = 38.20, SD_{age} = 12.80; 150 female, 104 male, 3 other). Participants received monetary compensation.

Stimuli. We used 40 male and 40 female faces with neutral expressions from the Chicago Face Database (Ma et al., 2015). To avoid interactions due to racial stereotypes, we used only White faces, and each participant rated 30 unique faces of only one gender (randomized across participants). For each participant, we randomly selected 30 faces from the pool of 40 faces. We used the 14 traits from Oosterhof and Todorov (2008): *aggressive, attractive, caring, confident, dominant, emotionally stable, intelligent, mean, responsible, sociable, threatening, trustworthy, unhappy,* and *weird*.

Procedure. First, we assigned participants to one of two groups using a classic minimal group paradigm procedure (Tajfel et al., 1971; Hong & Ratner, 2021). Participants were told that we were interested in how facial features relate to artistic preference. They were then told that that they would make judgments about 30 faces of people who have different artistic preferences on five different traits. Participants first completed a test to determine their own artistic preferences. In this task, they viewed 12 pairs of paintings by modern European artists, Paul Klee and Wassily Kandinsky, and chose whichever painting they liked better on a given trial. On each trial, one of the paintings was by Kandinsky and the other one was by Klee. The location of each painting did not correspond to the painter, and the signature of the painter was hidden from each painting to prevent participants from choosing based on the painter's name. At the end of the test, the computer program provided predetermined, bogus feedback (randomized across participants), indicating that each participant preferred paintings by either Kandinsky or Klee.

Face ratings. Following group assignment, participants rated 15 ingroup faces and 15 outgroup faces on 5 traits randomly selected from the total list of 14 traits. We presented each person's artistic preference along with their photograph (e.g., "How trustworthy is this *Kandinsky* person?"). Participants responded using a 7-point Likert scale (1 – not at all, 7 – very much). If the target person shared the same group membership as the participant (e.g., Kandinsky person rating Kandinsky person), then the participant was making judgments about an ingroup face; if they had different group memberships (e.g., Kandinsky person rating Klee person), then the participant was making judgments about an outgroup face. Participants rated each trait in a block before moving onto the next trait. The order of trait presentation was randomized across participants, and the order of face presentation was randomized across participants and across blocks. These ratings allowed us to generate a 5×5 similarity matrix in face ratings (within a broader 14×14 similarity matrix across participants) that captures the pairwise similarities (i.e., Pearson correlations) in face ratings among the 5 traits evaluated (see Figure 1).

Stereotype ratings. Next, participants rated stereotype associations for each of the two minimal groups (i.e., their conceptual associations about the two groups without any face stimuli involved). Participants were asked to report how they thought a person's artistic preference related to their personality. They rated stereotype associations for every possible pair of the same 5 traits from the face rating task, and they did so bidirectionally (e.g., "How likely is a trustworthy *Kandinsky* person to be also responsible?" and "How likely is a responsible *Kandinsky* person to be also trustworthy?" on a 7-point Likert scale (1 – not at all, 7 – very much). These ratings allowed us to generate a 5×5 similarity matrix in stereotype ratings (within a broader 14×14 similarity matrix across participants) that captures the pairwise similarities (i.e.,

average of the two bidirectionally assessed ratings) among the 5 traits evaluated (see Figure 1). This could then be directly compared to that using the face ratings.

Lastly, participants were asked about their agreement with the group assignment, their collective identification with their group, and any familiarity with Kandinsky and Klee and their paintings.

Results

The manipulation checks were successful, and prior familiarity with Kandinsky and Klee was low. We observed clear group differences in face ratings in favor of ingroup (Supplemental Analyses S2-S3).

Relationship between facial impression space and stereotype space. First, we examined whether each perceiver's unique stereotype associations predicted their own unique facial impression space for each perceiver by using a similar multilevel RSA approach (Xie et al., 2021). First, we created a 5×5 face rating correlation matrix for ingroup and outgroup, and Fisher-z transformed the correlations to compare across ingroup and outgroup. This resulted in 10 unique trait-pair correlations (e.g., trustworthy-attractive) per group per participant. These trait-pair correlations were then joined by corresponding trait-pair stereotype associations (with each being the average of the two bidirectionally assessed ratings). Our aim was to assess whether differences between ingroup and outgroup in stereotype space predicted corresponding differences in facial impression space. To isolate these group-based differences in trait-pair similarities, we subtracted outgroup trait-pair similarity indices (i.e., both face correlation and stereotype association) from corresponding ingroup trait-pair similarity indices for each participant. We used a multilevel approach to predict [ingroup - outgroup] facial impression space from [ingroup - outgroup] stereotype space (i.e., pairwise stereotype association) (see

Figure 1). The model allowed for random intercepts and slopes for participant and random intercepts for trait-pair. We found that differences between ingroup and outgroup in perceivers' stereotype space predicted corresponding differences in facial impression space, b=.04, β=.10, z=4.36, 95%CI[.02,.06], p<.001 (Figure 1).

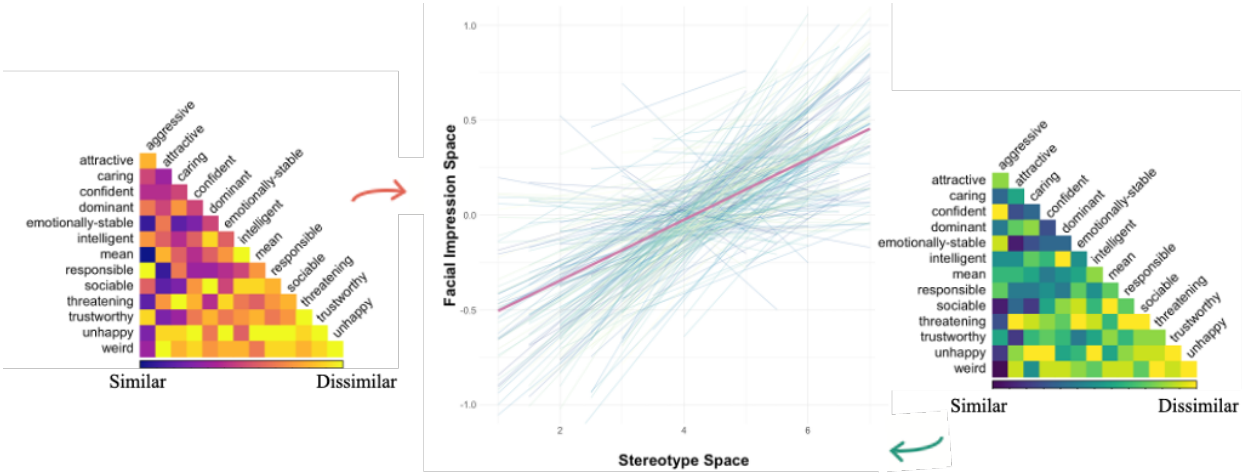

Figure 1. The relationship between stereotype space (right) and facial impression space (left) for each participant in Study 1. Each light blue line represents a regression line for each participant, and the red line represents an average regression line across all participants. Although the analysis involved Fisher's z transformation and subtraction between ingroup and outgroup, the relationship between stereotype space and facial impression space using untransformed data are depicted here for illustrative purposes. See Supplemental Table S2 for individual trait pair correlation/similarity scores for ingroup and outgroup.

Differences in core dimensions across group boundaries. After showing that the overall structures of facial impressions and stereotypes are different across group boundaries, we sought to better characterize the differences. We first reduced the dimensionality of our data to identify underlying dimensions of facial impressions and stereotypes. We used multidimensional scaling (MDS) to reduce the dimensionality of our data. Although our face rating data are suitable for more conventional dimensionality reduction techniques such as PCA (Oosterhof & Todorov, 2008; Jones et al., 2021), the stereotype association data, as similarity ratings, are not. As we

wanted to keep our analyses consistent and comparable to each other, we used MDS for the face ratings data as well.

Because each participant answered stereotype association between a given pair bidirectionally, we first added the two scores for each pair of traits, which resulted in scores ranging from 2 (very dissimilar) to 14 (very similar). We then recoded our stereotype association scores so that a higher score means more dissimilarity between two traits and constructed a 14×14 dissimilarity matrix (1 – identical traits in a dissimilarity matrix, 2 – very similar, 14 – very dissimilar), because MDS requires distance (dissimilarity) as input data. Based on previous research showing two central dimensions of face and person perception (Fiske et al., 2002; Oosterhof & Todorov, 2008), we used a two-dimensional solution. We then identified two clusters of traits by using k-means clustering analysis on each trait's scores on the two dimensions. K-means clustering is an unsupervised machine learning algorithm that uses an iterative refinement technique to find *k* number of clusters with the least amount of combined within-cluster variance. Each cluster is defined as data points with the least squared Euclidean distances to each other. Because we expected two core dimensions underlying person perception (i.e., trustworthiness and dominance), we used a 2-cluster solution. Figure 2A shows two clusters reflecting trustworthiness and dominance, which are generally consistent with past work (Oosterhof & Todorov, 2008; Jones et al., 2021).

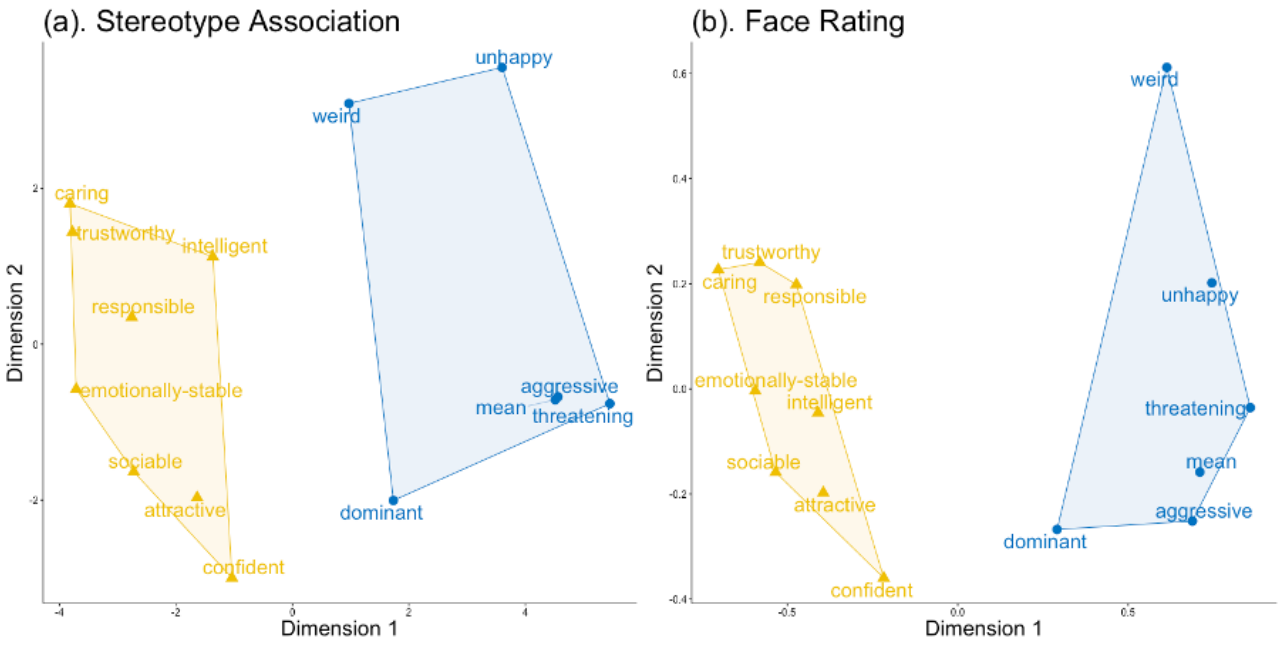

Figure 2. Multidimensional scaling results of Study 1 indicate two clusters of traits based on pairwise trait dissimilarities for (a) stereotype associations and (b) face ratings. For both ratings, the first cluster included attractive, caring, confident, emotionally stable, intelligent, responsible, sociable, and trustworthy (i.e., the trustworthiness cluster – yellow), whereas the second cluster included aggressive, dominant, mean, threatening, unhappy, and weird (i.e., the dominance cluster - blue),

For face ratings, we converted each pairwise correlation into distance by subtracting the correlation coefficient from 1 (1–r), as MDS requires dissimilarity as data input. We then used k-means clustering to identify 2 clusters of traits. We found identical clusters of traits with face ratings data as with stereotype association data (see Figure 2b). Independent raters confirmed that the two clusters correctly reflected trustworthiness and dominance (Supplemental Analysis S4).

Next, we computed average dissimilarity values between traits across the two clusters (e.g., caring-aggressive) for both stereotype associations and face ratings, separately for ingroup and outgroup. Because each participant rated 5 randomly selected traits out of the total 14, not all participants had the same number of cross-cluster dissimilarity scores. On average, each

participant had 5.21 cross-cluster dissimilarity scores (SD = 1.27), each for ingroup and outgroup. The number of cross-cluster dissimilarity scores did not differ between participants who were assigned to the Kandinsky group vs. Klee group, $t(255) = .94$, $p = .35$, 95% CI [-.16, .46]. Six participants rated all 5 traits within the same cluster, and thus were excluded from this specific analysis.

We used these cross-cluster dissimilarity data to test the critical question of whether targets' ingroup/outgroup status shifts the relationship between trustworthiness and dominance. We compared the mean-level cross-cluster dissimilarity scores between ingroup and outgroup using paired t-tests. For stereotype space, we found that traits in the trustworthiness cluster were more dissimilar to traits in the dominance cluster for outgroup (M = 8.31, SD = 1.97) relative to ingroup (M = 8.06, SD = 1.97), $t(249) = 3.50$, $p < .001$, 95% CI [.11, .40], Cohen's d = .22 (Figure 3). For facial impression space, we found that traits in the trustworthiness cluster were more negatively correlated with traits in the dominance cluster for outgroup (M=-.30, SD=.33) relative to ingroup (M = -.25, SD = .33), $t(249) = 3.20$, $p = .002$, 95% CI [.02, .09], Cohen's d = .20.

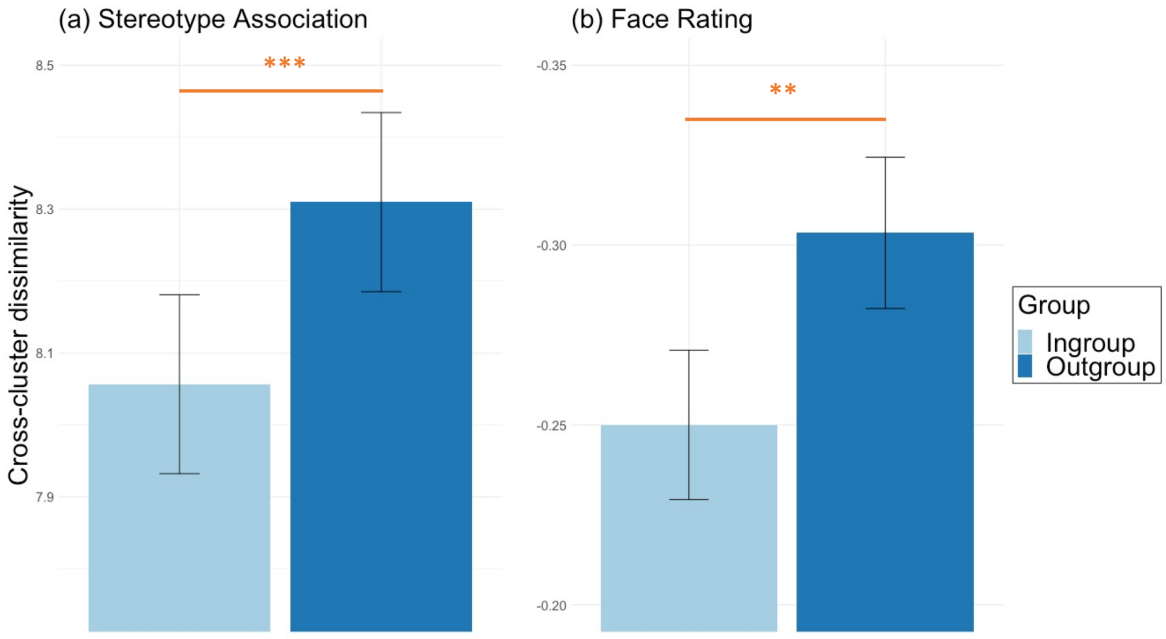

Figure 3. Bar plots comparing mean cross-cluster trait dissimilarity between ingroup and outgroup for (a) stereotype association and (b) face rating. Significance code: ** <.01 *** < .001

Additionally, across participants, we found a significant Spearman rank-order correlation between averaged [ingroup – outgroup] cross-cluster face rating dissimilarities (i.e., ingroup minus outgroup) and corresponding [ingroup – outgroup] cross-cluster stereotype dissimilarities, $\rho(248) = .17$, 95% CI = [.05, .29], $p = .007$ (Figure 4). This corroborates the finding of a significant relationship between unique variations in stereotype space predicting corresponding variations in facial impression space (Figure 1).

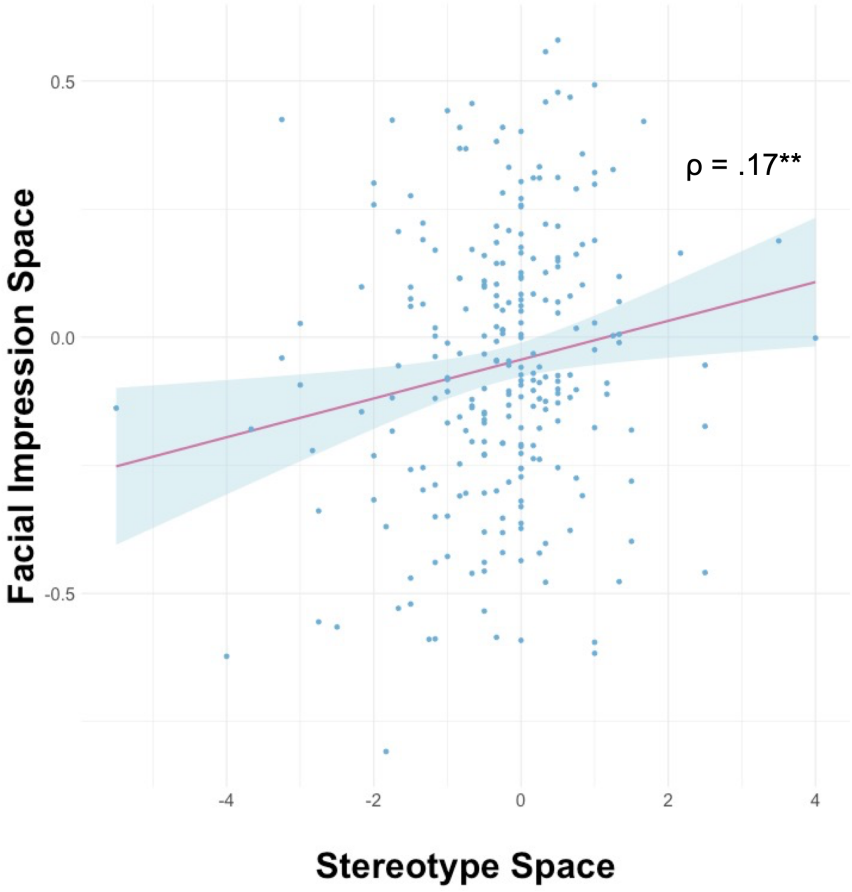

Figure 4. A Spearman rank-order correlation across participants between [ingroup – outgroup] cross-cluster face rating dissimilarities and corresponding [ingroup – outgroup] cross-cluster stereotype dissimilarities in Study 1. The extent to which trustworthiness and dominance components became more negatively related for outgroup relative to ingroup in stereotype associations predicted a corresponding change in a participant's face ratings. ** < .01

Supplemental Analysis S5 casts doubt on the possibility that these effects were driven by mere ingroup favoritism, as the shifts in core dimensions were unrelated to ingroup favoritism. Moreover, there was no evidence that these effects were limited to only one direction of influence, with trustworthiness and dominance impacting each other equally (Supplemental Analysis S6).

Discussion

In Study 1, we found that group-induced changes in conceptual associations were related to corresponding changes in how ingroup and outgroup faces were evaluated. Further, we found that such differences in stereotype and facial impression structures were due to differences in the relationship between trustworthiness and dominance. Trustworthiness and dominance became relatively more negatively related for the outgroup relative to ingroup.

Study 2

Study 2 aimed to corroborate the differential relationships between trustworthiness and dominance across ingroup and outgroup by manipulating rather than measuring trait-related facial appearance. We manipulated one trait dimension (e.g., trustworthiness) and examined its effects on the other dimension (e.g., dominance). Our critical prediction was that trustworthiness and dominance would become relatively more negatively related when evaluating outgroup relative to ingroup, which would be driven by different conceptual stereotype associations.

Methods

Participants. We recruited 300 White participants from Prolific to participate in an online study about making social judgments. Same as Study 1, we recruited White participants only to control for any effects of preexisting racial stereotypes and prejudices. We again excluded participants who did not have any variability in their responses or failed attention checks. After exclusion, our final sample size was 256 participants (M_{age}=37.34, SD_{age}=12.73; 119 female, 127 male, 7 other, 3 undisclosed). Participants received monetary compensation.

Stimuli. We used 16 White male face identities from the Basel Face Database (Walker et al., 2018). White faces were again used to avoid confounds related to gender and racial stereotypes. We also used only male faces, because there were not enough female faces in the database to make the number of stimuli equal across two genders. Each identity was

independently manipulated on the communion and the agency dimensions, which are effectively identical to the trustworthiness and dominance dimensions, respectively (Chua & Freeman, 2021; Walker et al., 2018). Thus, each face identity had 4 variations: high trustworthy, low trustworthy, high dominance, and low dominance (Figure 5). We labeled 8 face identities as ingroup and 8 face identities as outgroup, randomized across participants.

Procedure. The procedure was nearly identical to that of Study 1 except for the number and type of stimuli used. After minimal group assignment, participants made judgments of trustworthiness and dominance in separate blocks (block order counterbalanced). In each block, 16 unique face identities were presented, 8 of which were labeled as ingroup and the other 8 labeled as outgroup. Each identity was presented in low, original, and high variants. In each block, the facial dimension that varied was always the opposite of the trait dimension being assessed, and thus putatively irrelevant for the judgment (i.e., participants judged trustworthiness while facial dominance was manipulated, and vice versa, on a 7-point Likert scale: 1–not at all, 7–very much). The order of face presentation was randomized across participants and across blocks. Participants then rated stereotype associations between trustworthiness and dominance traits for each of the two minimal groups (e.g., “How likely is a trustworthy Kandinsky person to be also dominant?”), bidirectionally, on a 7-point Likert scale. Lastly, as in Study 1, participants answered manipulation check questions.

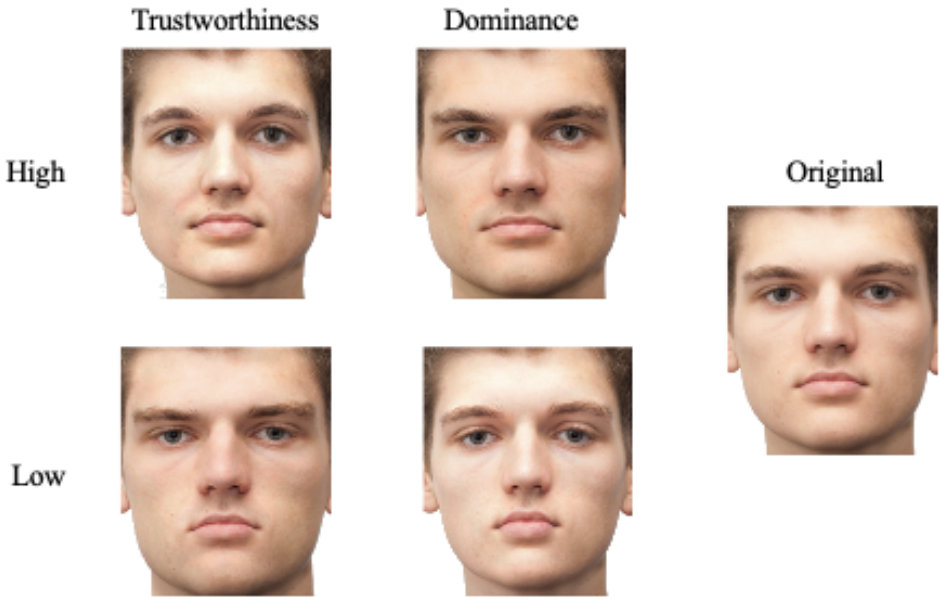

Figure 5. Example of an individual face identity manipulated to appear low or high in trustworthiness or dominance. Each identity was presented in low (-3 SD), original, and high (+3 SD) variants

Results

The manipulation checks were successful, and prior familiarity with Kandinsky and Klee was low (Supplemental Analysis S7).

First, we examined stereotype associations of trustworthiness and dominance. Consistent with Study 1, we added the two directional stereotype ratings for each participant, resulting in scores ranging from 2 (very dissimilar) to 14 (very similar), and recoded them so that a higher score means more dissimilarity. A paired t-test to showed that participants believed trustworthiness and dominance were more dissimilar for outgroup ($M = 8.34$, $SD = 2.15$) than ingroup ($M = 7.40$, $SD = .12$), $t(255) = 6.77$, 95% CI [.67, 1.20], $p < .001$, Cohen's $d=.42$.

Next, we used a linear mixed-effects model to predict trustworthiness/dominance face judgments from the face variant presented (low, original, or high variant of the alternate trait dimension), faces' group membership (ingroup, outgroup), and their interaction. In this model,

the manipulated level of stimuli was centered to show overall effects of group membership. The main effect of face variant was significant, indicating an overall negative relationship between trustworthiness and dominance, b = - .27, β = -.30, z = 22.31, 95% CI[-.29, -.24], p < .001. There was also a significant main effect of group membership, indicating that on average ingroup faces were judged more trustworthy and dominant than outgroup faces regardless of the manipulation, b = .15, β = .11, z = 4.70, 95% CI[.09, .22], p < .001. There was a significant interaction, b = .05, β = .04, z = 2.23, 95% CI[.01, .10], p = .03, which arose due to a more negative relationship between the two traits when judging outgroup faces (simple b = -.29, β = -.33, z = 17.39, 95% CI[-.33, -.26], p < .001) than ingroup faces (simple b = -.24, β = -.27, z = 14.23, 95% CI[-.27, -.21], p < .001). Additional analyses demonstrated that trustworthiness and dominance impacted each other symmetrically (Supplemental Analysis S6).

***Mediational analysis*.** Lastly, we conducted a mediational analysis to test for the intermediary role of stereotype associations in the shift in facial impressions across ingroup and outgroup. Specifically, we tested the possibility that stereotype associations (mediator) may partly explain the effects of ingroup/outgroup status (independent variable) on the relationship between trustworthiness and dominance dimensions (i.e., the effect of the trustworthiness/dominance face variants on judgments of the alternate dimension). For each participant we computed the correlation coefficient between low/original/high [-1,0,1] facial variants and judgments of the alternate trait (dependent variable), separately for ingroup and outgroup members [0,1]. For brevity, we refer to the correlation between facial variants and judgments of the alternate trait as "face judgments" below.

As expected, ingroup/outgroup status was significantly related to both face judgments, b = -.09, β = -.08, z = 2.51, 95% CI=[-.17, -.02], p = .01, and stereotype associations, b = .93, β

= .21, z = 6.73, 95% CI = [.66, 1.21], p < .001. More importantly, the relationship between stereotype associations and face judgments remained significant after controlling for ingroup/outgroup status, b = -.03, β = -.10, z = 2.26, 95% CI=[-.05, -.00], p = .02, whereas the relationship between group membership and face judgments was no longer significant, b = -.07, β = -.05, z = 1.70, 95% CI=[-.15, .01], p = .09, indicating evidence for full mediation. Finally, a 10,000-iteration bootstrapping analysis demonstrated a significant indirect effect, indicating that stereotype associations fully explained the relationship between group membership and face judgments, b = -.03, 95% CI=[-.05, -.00], p=.03 (Figure 6).

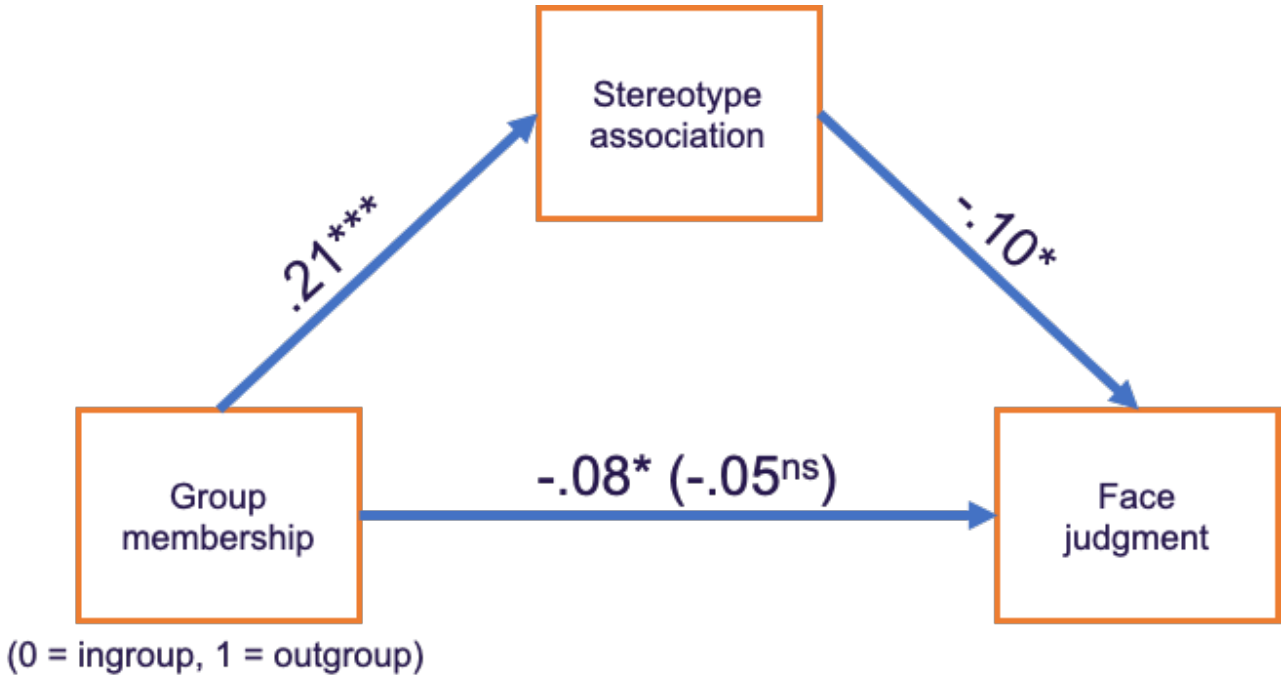

Figure 6. The model shows the effects of group membership (ingroup = 0, outgroup 1) on face judgments as mediated by stereotype associations. Values indicate standardized regression coefficients (β). Significance code: ns >.05 * <.05 *** < .001

Discussion

By manipulating trustworthiness and dominance of face stimuli and assessing their effects on the alternate trait dimension, we found converging evidence that these core dimensions

of facial impressions are differentially related across group boundaries. Moreover, we found that a difference in stereotypical beliefs across ingroup and outgroup drove these core dimensions to become differentially related in face judgments.

General Discussion

In two studies, we demonstrated shifts in core dimensions of facial impressions across minimal ingroup and outgroup. In Study 1, we first showed that there are variations in facial impression structures across group boundaries, which were explained by a shift in the two core dimensions, trustworthiness and dominance. Trustworthiness and dominance were stereotypically believed to be relatively more negatively related for outgroup than ingroup, and this predicted similar patterns in face judgments. In Study 2, we showed that systematically manipulating trait-related facial appearance (i.e., trustworthiness and dominance) resulted in perceptions of the alternate dimension that are different between ingroup and outgroup. We also demonstrated that a stronger stereotypic belief that trustworthiness ≈ dominance for ingroup than outgroup mediated the shifts in facial impression dimensions. Overall, these findings suggest that core dimensions of facial impression are flexible and can shift in intergroup contexts. These shifts occurred above and beyond mere favoritism, providing insight into our understanding of intergroup bias and perception.

It has long been suggested that the two core dimensions underlying facial impression structures, trustworthiness and dominance, are independent and relatively fixed and universal given their functionally adaptive nature (Jones et al., 2021; Oosterhof & Todorov, 2008). However, a growing body of research has demonstrated meaningful variability across individuals and cultures (Stolier et al., 2018). The present results add to this mounting evidence of a dynamic rather than fixed architecture for facial impression that is strongly context dependent (e.g., Oh et

al., 2022), varying across different social cognitive factors, such as targets' ingroup/outgroup status. Specifically, our work provides converging evidence that changes in individuals' idiosyncratic stereotypic beliefs about groups predict corresponding changes across groups (e.g., race and gender in Xie et al., 2021). Our work expands on this finding by isolating the effects of ingroup/outgroup status using the minimal group paradigm. By doing so, our work advances our understanding of intergroup perception by showing that the effects of group membership on facial impression and stereotype structures go beyond mere ingroup favoritism. While it is widely known that ingroup members are perceived more favorably than outgroup members across different traits (e.g., Ratner et al., 2014), here we showed that ingroup/outgroup distinction influenced the structure in which different traits were perceived.

We argued that trustworthiness and dominance become relatively more negatively related for outgroup members due to dominance being registered as an intergroup threat. Outgroup members who are more dominant would be better able to enact bad intentions, such as taking away one's resources (Esses et al., 1993), and consequently be perceived as more untrustworthy. Ingroup members, however, are expected to provide support and resources (Brewer, 2004), and thus those who are more dominant would be better able to enact good intentions would be perceived as competent (Anderson & Kilduff, 2009) and consequently more trustworthy (Oliveira et al., 2019). This leads to shifts in perceivers' conceptual knowledge about the ingroup and outgroup and, in turn, shifts in facial impression structures. Indeed, it has long been known that threatening faces are judged more dominant and less trustworthy, while non-threatening faces judged more competent and more trustworthy (Todorov et al., 2008). The overall effects we observed are also consistent with findings of motivated perception and cognition (Balcetis & Dunning, 2006; Hughes & Zaki, 2015). For instance, perceivers more strongly process goal-

consistent information (e.g., ingroup = good) while often failing to appropriately process goal-inconsistent information (Hughes et al., 2017). Accordingly, it is possible that trustworthiness and dominance being relatively more negatively related for outgroup than ingroup reflects similar motivational processes that change both conceptual associations about ingroup and outgroup and, consequently, facial impression structures. While motivational processes may be at the heart of these shifts, our results suggest that a proximal contributor to changes in facial impressions structure is conceptual associations about ingroup and outgroup.

Although our work provides a first demonstration of shifts in core dimensions of facial impressions across group boundaries, there are several limitations. First, we only used White face stimuli and recruited White participants to limit the effects of preexisting stereotypes and prejudices about real-world groups. Furthermore, we used novel groups using the minimal group paradigm, which may limit the generalizability of our findings. That said, these findings provide a foundation on which ingroup/outgroup distinctions may have in forming impressions of others' faces. We would expect such intergroup effects to be more nuanced and complex in the context of real-world groups such as racial or political groups due to integrative processing of pre-existing stereotypes and attitudes (Freeman et al., 2020; Kunda & Thagard, 1996). An important next step of this research is to examine the effects of ingroup/outgroup distinctions in the presence of real-world groups. Second, traits in the trustworthiness and dominance clusters identified in Study 1 differed significantly on valence (Supplemental Analysis S4), raising the possibility that the shifts in facial impression structures across group boundaries may in part be due to changes in the relationship between positive and negative traits. Although this does not discount our findings that overall facial impression structures shift across ingroup and outgroup, future research could more clearly delineate the mechanisms behind such shifts by attempting to

distinguish valence from the core dimensions identified in our research. Lastly, it is possible that ingroup/outgroup status could change person perception dimensions beyond the two-dimensional model, such as across three (Sutherland et al., 2013) or four dimensions (Lin et al., 2021). These questions could be explored by future research.

In conclusion, the present work provides evidence for the role of group membership in shifting facial impression structures. Our findings not only further show that core dimensions of facial impressions are more flexible and context-dependent than typically appreciated, but also show that intergroup contexts exert fundamental changes in the structure of facial impressions that go beyond ingroup favoritism and outgroup derogation.

References

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology*, *96*, 491–503.

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*, 612–625.

Beaupré, M. G., & Hess, U. (2003). In my mind, we all smile: A case of in-group favoritism. *Journal of Experimental Social Psychology*, *39*(4), 371–377.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706–712.

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, *62*, 728–738.

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

Cogsdill, E., Todorov, A., Spelke, E., & Banaji, M. (2014). Inferring Character From Faces: A Developmental Study. *Psychological Science*, *25*(5), 1132–1139.

Collova, J. R., Sutherland, C. A. M., & Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of Personality and Social Psychology*, *117*, 900–924.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R.

(2009). Stereotype content model across cultures: Towards universal similarities and some differences. *The British Journal of Social Psychology / the British Psychological Society*, *48*(0 1), 1–33.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and Credit: The Role of Appearance in Peer-to-peer Lending. *The Review of Financial Studies*, *25*(8), 2455–2484.

Dunham, Y. (2011). An angry = Outgroup effect. *Journal of Experimental Social Psychology*, *47*(3), 668–671.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Chapter 7 - Values, Stereotypes, and Emotions as Determinants of Intergroup Attitudes. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, Cognition and Stereotyping* (pp. 137–166). Academic Press.

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

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology*, *61*, 237–287.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852.

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), e12431.

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529.

Hong, Y., Mayes, M. S., Munasinghe, A. P., & Ratner, K. G. (2022). Scrutinizing Whether Mere Group Membership Influences the N170 Response to Faces: Results from Two Preregistered Event-Related Potential Studies. *Journal of Cognitive Neuroscience*, *34*(11), 1999–2015.

Hong, Y., & Ratner, K. G. (2021). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology*, *120*, 576–600.

Hugenberg, K. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168.

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science*, *33*(6), 1173–1181.

Hughes, B. L., & Zaki, J. (2015). The neuroscience of motivated cognition. *Trends in Cognitive Sciences*, *19*(2), 62–64.

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience*, *12*(1), 49–60.

Hutchings, R. J., Simpson, A. J., Sherman, J. W., & Todd, A. R. (2021). Perspective taking reduces intergroup bias in visual representations of faces. *Cognition*, *214*, 104808.

Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology*, *90*, 104004.

Jones, B. C. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*, 13.

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, *111*(5), 655–664.

Lazerus, T., Ingbretsen, Z. A., Stolier, R. M., Freeman, J. B., & Cikara, M. (2016). Positivity bias in judging ingroup members' emotional expressions. *Emotion (Washington, D.C.)*, *16*(8), 1117–1125.

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications, 12*(1), 1–15.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Nature Human Behavior, 4*, 287-293.

Oh, D., Martin, J. D., & Freeman, J. B. (2022). Personality Across World Regions Predicts Variability in the Structure of Face Impressions. *Psychological Science*, *33*(8), 1240–1256.

Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, *49*(5), 888–902.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Ratner, K. G., & Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology*, *49*(2), 298–301.

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., 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*(6), 897–911.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, *104*(3), 409–426.

Stolier, R.M., Hehman, E.A., & Freeman, J.B. (2018). A Dynamic Structure of Social Trait Space. *Trends in Cognitive Sciences, 22*(3), 197-200.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371.

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, *115*(37), 9210–9215.

Sutherland, C. A. M., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences*, *117*(19), 10218–10224.

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*(4), 521–537.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*(1), 105–118.

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology (London, England: 1953)*, *106*(2), 186–208.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of Competence from Faces Predict Election Outcomes. *Science*, *308*(5728), 1623–1626.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460.

Walker, M., Schönborn, S., Greifeneder, R., & Vetter, T. (2018). The Basel Face Database: A validated set of photographs reflecting systematic differences in Big Two and Big Five personality dimensions. *PLOS ONE*, *13*(3), e0193190.

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, *26*(8), 1325–1331.

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

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial Impressions Are Predicted by the Structure of Group Stereotypes. *Psychological Science*, *32*(12), 1979–1993.