Social Cognition, Vol. 41, No. 6, 2023, pp. 562–578

FACIAL STEREOTYPES OF COMPETENCE (NOT TRUSTWORTHINESS OR DOMINANCE) MOST RESEMBLE FACIAL STEREOTYPES OF GROUP MEMBERSHIP

Youngki Hong
Columbia University and University of California, Santa Barbara

Megan Reed and Kyle G. Ratner
University of California, Santa Barbara

Previous research shows that perceivers have distinct mental representations of ingroups and outgroups even when groups are novel and not defined by physical attributes. Here, we leverage the minimal group paradigm, the reverse correlation method, and machine learning to parse the visual ingredients of group membership. In Study 1, we found that ingroup faces are trusted more than outgroup faces and that facial stereotypes of trustworthiness resemble those of the ingroup/outgroup distinction. However, in Study 2 we showed that such facial stereotypes of group membership resembled those of competence more than trustworthiness and dominance. Together, these findings suggest that even though trustworthiness is an important visual ingredient of the ingroup/outgroup distinction, people may rely on facial cues indicating competence the most to guide their visualization of novel ingroup and outgroup members, highlighting the nuanced nature of ingroup bias in face processing.

Keywords: faces, minimal group paradigm, reverse correlation, machine learning

Recent applications of reverse correlation image classification methods from visual cognition to social psychology have proven useful for understanding social category and person identity representation related to physical attributes, including racial groups (Dotsch et al., 2008; Imhoff et al., 2011) and familiar individuals (Mangini & Biederman, 2004; Oh et al., 2021; Young et al., 2014). In these cases, the target of representation is one that the perceivers have encountered previously, so

Address correspondence to Youngki Hong, Department of Psychology, Columbia University, New York, NY 10027. E-mail: youngkih41@gmail.com

the representations revealed by the reverse correlation procedure can be explained by participants populating their mental images with exemplar or prototype information retrieved from memory (Nosofsky & Zaki, 2002). However, other research suggests that visual representations might not always require such memory traces. Specifically, people can visually imagine the appearance of fellow group members when these groups are completely novel and not defined by physical appearance (Hong & Ratner, 2021; Hutchings et al., 2021; Ratner et al., 2014). These studies suggest that people's mental image of novel ingroup faces elicits more desirable trait impressions (e.g., trustworthy) than their outgroup counterparts. However, it is still unclear whether people rely on facial cues related to trust or other traits critical to face processing, such as competence or dominance, to guide their visualizations of novel ingroup and outgroup members.

There are a lot of reasons to predict that differences in representations of ingroup and outgroup faces are driven by facial stereotypes (Chua & Freeman, 2021) specific to trust. Trustworthiness is considered a central dimension of face perception for signaling whether someone is perceived to have good or bad intentions (Oosterhof & Todorov, 2008). Perceived trustworthiness is also related to many important social outcomes, including financial decision making, personnel selection, and criminal sentencing (Duarte et al., 2012; Olivola et al., 2014; Wilson & Rule, 2015). In addition, people trust ingroup members more than outgroup members (Foddy et al., 2009; Tanis & Postmes, 2005). People may do so even in novel group situations because they associate more positive qualities with ingroup members than with outgroup members (Brewer & Silver, 1978). Recent studies show that such a bias extends to face processing. For example, people visualize faces of ingroup members as more trustworthy-looking than faces of outgroup members (Ratner et al., 2014) and accept more trustworthy-looking faces into their ingroup (Tracy et al., 2020). Thus, given the central role that trustworthiness plays in face perception and in intergroup perception, it is plausible that people populate their mental image of novel ingroup members with features that convey trustworthiness more than they do for outgroup members. This could be accomplished because judgments of trustworthiness from faces are strongly related to a face's physical resemblance to emotional expressions of happiness, such as joy indicated by an upturned mouth (Kleisner et al., 2013; Oosterhof & Todorov, 2008; Zebrowitz et al., 2003).

However, many consequential judgments about ingroup others that on the surface could be attributed to trust have been most directly linked to inferences of competence, not trust. For instance, facial competence is related to election results and leadership attainment (Antonakis & Eubanks, 2017; Todorov et al., 2005). Relatedly, according to the Stereotype Content Model, ingroups are often stereotyped as competent (Cuddy et al., 2009; Fiske et al., 2002, 2007), indicating that competence may be an important social cue that distinguishes ingroups from outgroups. Critical to the current investigation, competence seems to be related to trustworthiness (Oliveira et al., 2019) and also to dominance (Anderson & Kilduff, 2009). Dominance has been conceptualized as the other central dimension in face perception (in addition to trustworthiness) and signals someone's perceived ability to

enact good or bad intentions (Oosterhof & Todorov, 2008). Although dominance is often associated with negative attributes such as aggression, low intelligence, and untrustworthiness (Carré et al., 2009; Stirrat & Perrett, 2010), dominance is sometimes preferred in ingroup members (Hehman et al., 2015) and may be perceived as competence in these members (Anderson & Kilduff, 2009).

In the current research, we started with the prediction that visual signals of trustworthiness differentiate facial stereotypes of ingroups versus outgroups. To test this idea, in Study 1 we used publicly available reverse correlation participant-level classification images (CIs) of novel ingroup and outgroup members (Hong & Ratner, 2021) in an economic trust game. We then used machine learning to test whether facial stereotypes of trustworthiness and group membership share any similarities. In Study 2, we used multiple regression to pit facial stereotypes of trustworthiness, dominance, and competence against each other in their contributions to facial stereotypes of group membership. Overall, we showed that although trustworthiness is an important cue that people use to guide their visualization of novel ingroup and outgroup faces (Study 1), unique facial stereotypes of competence (more than trustworthiness or dominance) resemble facial stereotypes of group membership (Study 2).

All data, study materials, and analysis scripts are publicly available at https://osf.io/8fzgj/.

STUDY 1

METHOD

The current research used publicly available CI data from Hong and Ratner (2021; https://osf.io/s9243/?view_only=92afae84a38548e8a9412e8353f30905) as stimuli. This stimuli set included two samples of participant-level CIs of novel ingroup and outgroup faces. The two samples were identical except for the version of the minimal group paradigm used (Tajfel et al., 1971). For information about how these CIs were generated, please see Studies 1 and 2 in Hong and Ratner (2021).

Part 1: Assessing Perceived Trustworthiness of Novel Ingroup and Outgroup Faces

Participants. To remove any confusion regarding the source of participant-level CIs, we use Sample 1 to refer to data from Study 1 and Sample 2 to refer to data from Study 2 of Hong and Ratner (2021). We recruited two samples of American college students to participate in an economic trust game designed to assess perceived trustworthiness of each participant-level CI.[1] The first sample played trust games with 362 participant-level CIs from Sample 1 (N = 108, M_{age} = 18.77,

1. While this version of the economic trust game successfully captured perceived trustworthiness of the CIs, there is no clear advantage to it over simple trustworthiness ratings, given that the participants played the games with artificial partners represented by the CIs and no stakes.

SD = 1.37; 67 female, 41 male). Racial and ethnic breakdown of this sample was 40 White, 32 Asian, 20 Latinx, 1 Black, 10 multiracial, and 5 other. Up to four participants were run simultaneously. The second sample played trust games with 200 participant-level CIs from Sample 2 (N = 148, M_{age} = 19.16, *SD* = 1.44; 94 female, 54 male). The racial and ethnic breakdown of this sample was 50 Asian, 38 Latinx, 38 White, 16 multiracial, 3 Black, 1 Pacific Islander/Hawaiian, 2 other, and 1 unidentified. We did not predetermine our sample size, but instead we ran as many participants as possible in a single 10-week academic quarter. All the analyses were conducted after data collection concluded.

Procedure. In this study, participants played an economic trust game with various interaction partners. The interaction partners were the participant-level CIs of novel ingroup and outgroup face images from Hong and Ratner (2021). We instructed participants to imagine that they had $10 on each trial and that they could choose either to keep this money or to share a certain amount with their interaction partners. On each interaction trial, participants made a choice to share a portion of $10 (i.e., $0, $2, $4, $6, $8, or $10). Participants were informed that any money they shared would be quadrupled and given to the interaction partner. The interaction partner would then have the option to return half of the sum to the participant who had shared the money. In this way, it was possible for the participant to make more money than if they had not shared. Participants simply indicated how much money they would like to share with each partner and did not receive any feedback. The amount of money shared was therefore indicative of the extent to which the participants trusted the interaction partners. A total of 108 participants played the trust game with 362 different partners from Sample 1, and 148 participants played the trust game with 200 different partners from Sample 2. The order of presentation of different CIs was randomized across participants.

Part 2: Assessing Facial Stereotypes of Trustworthiness and Group Membership and Their Similarity

Next, we used machine learning to classify each image as ingroup or outgroup and to predict the amount of money each image received in the trust game based on pixel intensity data. If the machine learning algorithm could successfully learn the association between pixel intensity data and each image's ingroup/outgroup status, as well as the amount of money received in the trust game, it would suggest that there are representational differences between ingroup and outgroup participant-level CIs and that perceived trustworthiness is reflected in each image. We then compared the importance of each variable (i.e., pixel) used in the algorithm for classifying between ingroup and outgroup and for predicting perceived trustworthiness for each sample.

First, we replicated the machine learning analyses from Hong and Ratner (2021) to classify participant-level CIs as either ingroup or outgroup, incorporating a few improvments. We (a) extracted faces from each CI using OpenFace's face

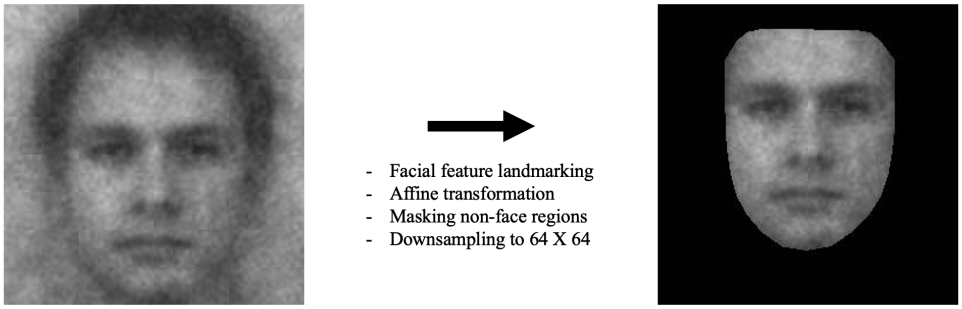

FIGURE 1. An example of face extraction of a participant-level CI using OpenFace (https://github.com/TadasBaltrusaitis/OpenFace).

extraction tool (see Figure 1; Amos et al., 2016); (b) applied an affine transformation so that each face's eyes, nose, and mouth appear in approximately the same location; (c) down-sampled pixel intensity data of each image from 512 × 512 to 64 × 64; (d) standardized the pixel intensity data; and (e) performed classification using support vector machines (SVM) with a linear kernel. We then used 10-fold cross-validation with our SVM model to minimize overfitting our data. Each fold yielded a training set (90% of the data) and a testing set (10% of the data), both evenly divided between ingroup and outgroup images. The SVM algorithm then learned the relationships between 64 × 64 pixel intensity data of each image and class labels (ingroup or outgroup) from the training set and classified images from the testing set that were not part of the training set for a given fold. We repeated this step 10 times until every instance of data was in both the training and testing sets at some point. We then computed accuracy scores by averaging classification accuracies across these 10 folds. Next, we used permutation tests to determine whether the accuracies of our machine learning algorithm significantly differed from chance (Ojala & Garriga, 2010). For each permutation, class labels (ingroup or outgroup) were randomly permuted for every image, removing any systematic relationship between pixel intensity data and class labels if there were any, followed by the classification steps described above. We repeated the same procedure 1,000 times (i.e., 1,000 permutation tests), allowing us to estimate the p value (i.e., the percentage of permutation tests that had higher accuracy than the accuracy with true labels).

For predicting perceived trustworthiness from each participant-level CI, we used a special type of SVM called support vector regression (SVR; Drucker et al., 1997) because perceived trustworthiness is a continuous outcome variable (i.e., the average amount of money each image received in the trust game). The SVR follows the same logic as the SVM and is thus better suited for analyzing data with high dimensionality (i.e., high number of predictors) such as images (e.g., a participant-level CI contains 64 × 64 = 4,096 predictor variables) compared to the ordinary

least-square linear regression. We used the SVR to predict the amount of money each participant-level CI received during the trust game based on pixel intensity data, following the similar steps described above, including 10-fold cross-validation and 1,000 permutation tests. Instead of classification accuracy, however, the model performance was measured by taking the mean absolute deviation (MAD; i.e., the average absolute value of predicted perceived trustworthiness minus actual perceived trustworthiness). Because smaller MAD values indicate better performance, we estimated the *p* value from the proportion of permutation MADs that were smaller than the true MAD.

Lastly, we performed feature selection using the variable ranking method (Guyon & Elisseeff, 2003) on the two SVMs for ingroup/outgroup and the two SVRs for predicting perceived trustworthiness. This method ranks feature importance based on *t* statistics (for regression) or the area under the ROC (receiver operating curve; for classification) associated with each variable (i.e., pixel). It indicates that features with higher values are more important than those with lower values in classifying between labels or predicting continuous outcome values. We constructed a variable importance matrix for each analysis and then overlaid them on the base image to visualize any clusters of regions on the face that were important for classifying between groups or predicting perceived trustworthiness (see Figure 2). Variable importance values were scaled to range between 0 and 100, making the analyses comparable across classification and prediction, with a greater number representing greater importance. Finally, we computed Pearson correlations between every pair of variable importance analyses to examine the similarity between facial stereotypes of perceived trustworthiness and group membership. If participants indeed relied on facial cues signaling trustworthiness to guide their decisions during the face categorization task, then the variables important for classifying between ingroup and outgroup images should be significantly correlated with the variables important for predicting the perceived trustworthiness of images.

RESULTS

Trust Game. We first averaged the amount of money each participant-level CI received in the trust game and conducted independent-samples *t* tests to examine whether participant-level CIs of novel ingroup faces received more money than outgroup faces. The results of Sample 1 showed that ingroup CIs received significantly more money ($M = 2.98$, $SD = .64$) than outgroup CIs ($M = 2.82$, $SD = .69$), $t(360) = 2.32$, $p = .02$, Cohen's $d = .24$. The results of Sample 2 replicated the finding: ingroup CIs received significantly more money ($M = 3.10$, $SD = .82$) than outgroup CIs ($M = 2.17$, $SD = .55$), $t(198) = 9.40$, $p < .001$, Cohen's $d = 1.33$. These results indicate that ingroup face images were more trustworthy-looking than their outgroup counterparts.

Based on sensitivity analyses conducted using G*Power (Faul et al., 2007), Sample 1 ($n = 362$) required a minimum effect size of Cohen's $d = .295$, and Sample 2

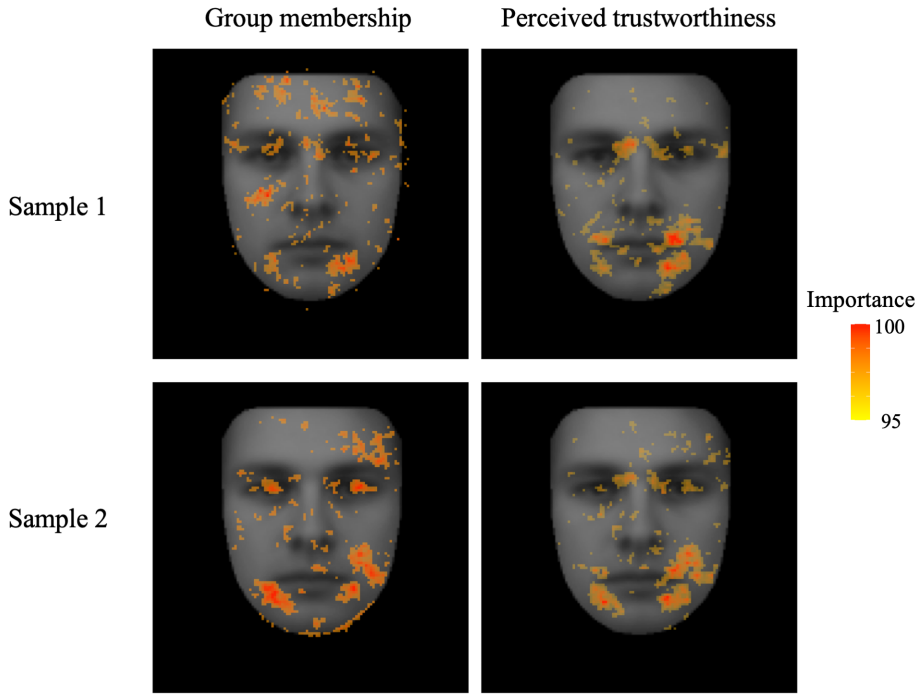

FIGURE 2. Variable importance maps of group membership classification and trustworthiness prediction. Note that only top 95th percentile of variables is shown.

(n = 200) required a minimum effect size of Cohen's d = .398, with a power of 80% and an alpha level of .05. This indicated that Sample 1 fell short of the minimum effect size requirement, whereas Sample 2 greatly exceeded the required minimum effect size.

Machine Learning. We classified between ingroup and outgroup images based on pixel intensity data of masked CIs better than chance (50%) for both samples (Sample 1 accuracy = 57.48%, p = .01; Sample 2 accuracy = 71.00%, p < .001). We predicted perceived trustworthiness of each CI significantly better than chance for both samples (Sample 1 MAD = .29, p < .001; Sample 2 MAD = .34, p < .001). Permutation test results are shown in Figure 3.

Similarity Analysis. Next, the correlation analysis showed that variable importance of the ingroup/outgroup classifications from the two samples was significantly correlated, $r(1745)$ = .18, p < .001. Not surprisingly, the variable importance of the perceived trustworthiness of the two samples was also significantly correlated, $r(1745)$ = .79, p < .001. More critically, we also found that variable importance for ingroup/outgroup classification and perceived trustworthiness prediction were significantly related for both Sample 1, $r(1745)$ = .17, p < .001 and Sample 2,

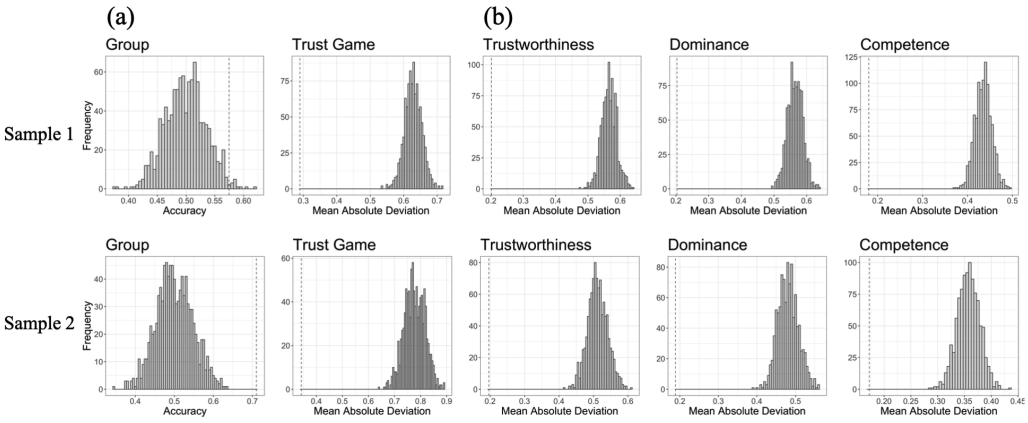

FIGURE 3. Permutation test results for (a) group membership classification, trust game prediction (Study 1), and (b) trait prediction (Study 2). The dotted lines indicate true accuracy/MAD scores.

$r(1745) = .66$, $p < .001$. All the correlation coefficients, p values, and confidence intervals are presented in Table 1. These results indicate that facial regions related to classifying ingroup and outgroup faces are similar to facial regions related to predicting perceived trustworthiness, which provides initial evidence that facial stereotypes of group membership are constructed in part with facial features related to trustworthiness.

STUDY 2

Although Study 1 demonstrated some commonality across facial stereotypes of group membership and trustworthiness, it is possible that facial stereotypes of group membership consist of multiple traits, and trustworthiness may not be at the core of the ingroup face representation. Furthermore, we cannot rule out the possibility that machine learning algorithms simply picked up on "signals" in the face images rather than clusters of regions meaningfully related to group membership or trustworthiness. In other words, representational differences, whether between ingroup and outgroup or trustworthy and untrustworthy faces, may exhibit similar regions of importance (e.g., eyes, mouth). This is particularly true because differences in the extent of holistic processing of faces in intergroup contexts (Hugenberg & Corneille, 2009) and trait impressions (Abbas & Duchaine, 2008) might be represented in similar facial regions related to both group membership and trustworthiness. Thus, in Study 2, we examined two additional traits, competence and dominance, both of which are important in face perception and intergroup perception.

TABLE 1. Correlation Matrix of Variable Importance Across Classifications and Regressions

Type	1	2	3
1. Group (Sample 1)			
2. Group (Sample 2)	.18***		
	[.13, .23]		
3. Perceived trustworthiness (Sample 1)	.17***	.36***	
	[.12, .21]	[.32, .40]	
4. Perceived trustworthiness (Sample 2)	.22***	.66***	.79***
	[.17, .26]	[.63, .68]	[.77, .80]

***$p < .001$

METHODS

Participants. We recruited two groups from Prolific to participate in an online study about how people make social judgments. We aimed to collect $n = 50$ per trait per sample based on similar studies examining trait impression differences (Ratner et al., 2014). The first group rated participant-level CIs from Sample 1 on one of three traits (trustworthiness, dominance, competence) ($N = 159$, $M_{age} = 42.23$, $SD = 15.26$; 86 female, 70 male, 3 other). Racial and ethnic breakdown of this sample was 101 White, 19 Latinx, 11 multiracial, 10 Asian, 10 Black, and 8 other. The trait breakdown of this sample was 54 trustworthiness, 54 dominance, and 51 competence. The second group rated participant-level CIs from Sample 2 ($N = 151$, $M_{age} = 38.49$, $SD = 11.51$; 58 female, 93 male). The racial and ethnic breakdown of this sample was 101 White, 19 Asian, 14 Latinx, 10 Black, 6 multiracial, and 1 other. The trait breakdown of this sample was 49 trustworthiness, 53 dominance, and 49 competence. All analyses were conducted after data collection concluded.

Procedure. In this study, participants rated participant-level CIs on one of three traits, trustworthiness, dominance, or competence (e.g., "How trustworthy is this person?"), using a 7-point Likert scale (1 = *not at all*, 7 = *very much*). A total of 159 participants rated all 362 CIs from Sample 1, and 151 participants rated all 200 CIs from Sample 2. The order of presentation of different CIs was randomized across participants.

Next, we used machine learning to predict trustworthiness, dominance, and competence ratings for each image using pixel intensity data. We used the same method used in Study 1 for predicting perceived trustworthiness, involving face extraction with OpenFace, SVR, cross-validation, permutation tests, and feature selection. For a detailed explanation of this method, see Part 2 of the Study 1 Procedure. To discern the unique contribution of each trait, we used multiple regression

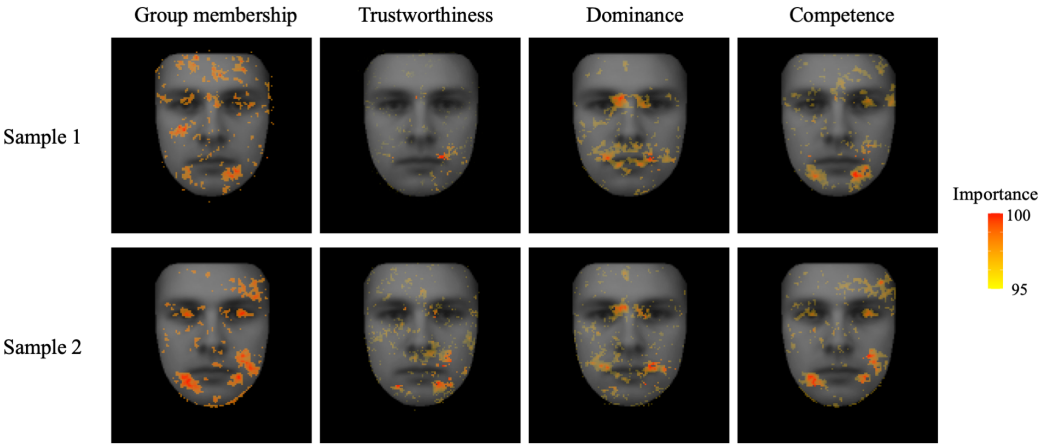

FIGURE 4. Variable importance maps of group membership classification and contrasted variable importance maps of trait prediction. Note that only top 95th percentile of variables is shown.

to predict variable importance for group membership classification. This involved a linear combination of variable importance scores for trustworthiness, dominance, and competence predictions. Before proceeding, we compared each variable's importance against the other traits, considering the high correlations among trait ratings (absolute $r > .8$). Any variable smaller than the corresponding variables of the other two traits was scored as 0 (of no importance), highlighting the unique contribution of a given trait (trustworthiness, dominance, or competence). We then incorporated these contrasted variable importance scores into the multiple regression models to predict facial stereotypes of group membership, illustrating the unique contributions of trustworthiness, dominance, and competence in relation to the variable importance for group membership in Figure 4.

RESULTS

Trait Rating. We first averaged ratings of each participant-level CI for each trait and ran independent-samples t tests to examine whether participant-level CIs of novel ingroup and outgroup faces elicited different trait impressions. For Sample 1, the results showed that ingroup CIs were seen as more trustworthy ($M = 3.50$, $SD = .55$) than outgroup CIs ($M = 3.37$, $SD = .63$), $t(360) = 2.11$, $p = .04$, Cohen's $d = .22$. Ingroup CIs were also perceived as more competent ($M = 4.07$, $SD = .41$) than outgroup CIs ($M = 3.92$, $SD = .47$), $t(360) = 3.11$, $p = .002$, Cohen's $d = .33$. There was no significant difference in dominance ratings of ingroup ($M = 4.22$, $SD = .57$) and outgroup ($M = 4.33$, $SD = .61$), $t(360) = 1.80$, $p = .07$, Cohen's $d = .19$.

For Sample 2, ingroup CIs were seen as more trustworthy ($M = 4.13$, $SD = .47$) than outgroup CIs ($M = 3.56$, $SD = .40$), $t(198) = 9.17$, $p < .001$, Cohen's $d = 1.30$, more competent (ingroup $M = 4.54$, $SD = .26$) than outgroup CIs ($M = 4.13$, $SD = .34$), $t(198) = 9.62$, $p < .001$, Cohen's $d = 1.36$, and less dominant (ingroup $M = 4.09$,

SD = .50) than outgroup CIs (*M* = 4.60, *SD* = .34), $t(198) = 8.41$, $p < .001$, Cohen's $d = 1.19$.

The same sensitivity power analyses reported in Study 1 applied to the current analyses: Sample 1 ($n = 362$) required a minimum effect size of Cohen's $d = .295$, and Sample 2 ($n = 200$) required a minimum effect size of Cohen's $d = .398$, with a power of 80% and an alpha level of .05. This indicated that only the competence ratings for Sample 1 met the minimum effect size requirement, whereas all three trait ratings for Sample 2 met the minimum effect size requirement.

Machine Learning. For both samples, we successfully predicted all three trait ratings of each CI significantly better than chance. This is indicated by true MAD values that are smaller than permutation MAD values (Sample 1 trustworthiness = .20, $p < .001$, dominance = .20, $p < .001$, competence = .18, $p < .001$; Sample 2 trustworthiness = .20, $p < .001$, dominance = .19, $p < .001$, competence = .17, $p < .001$). Permutation results are shown in Figure 3.

Multiple Regression. For Sample 1, multiple regression showed that unique variable importance of competence prediction is a significant predictor of variable importance of group membership classification, $\beta = .34$, $SE = .04$, $t(1743) = 14.27$, $p < .001$, and so was trustworthiness, $\beta = .05$, $SE = .05$, $t(1743) = 2.28$, $p = .02$. Dominance was not a significant predictor, $\beta = .01$, $SE = .03$, $t(1743) = .36$, $p = .72$. We conducted linear hypothesis testing to test whether trustworthiness and competence were significantly different from each other and found that competence predicted group membership significantly better than trustworthiness, $F(1, 1743) = 55.11$, $p < .001$. For Sample 2, all three measures of unique variable importance were significant predictors of variable importance of group membership classification: trustworthiness, $\beta = .32$, $SE = .03$, $t(1743) = 20.10$, $p < .001$; competence, $\beta = .80$, $SE = .02$, $t(1743) = 48.04$, $p < .001$; and dominance, $\beta = .28$, $SE = .02$, $t(1743) = 17.22$, $p < .001$. Linear hypothesis testing showed that competence was the best predictor of group membership [against trustworthiness $F(1, 1743) = 214.61$, $p < .001$; against dominance $F(1, 1743) = 475.72$, $p < .001$], followed by trustworthiness [against dominance $F(1, 1743) = 17.91$, $p < .001$] and dominance.

Lastly, we cross-validated our results by predicting variable importance of group membership classification from one sample using variable importance of trait predictions from the other sample. For Sample 1 group membership classification, all three traits from Sample 2 were significant predictors: trustworthiness, $\beta = .14$, $SE = .04$, $t(1743) = 5.96$, $p < .001$; competence, $\beta = .22$, $SE = .03$, $t(1743) = 9.01$, $p < .001$; and dominance, $\beta = .05$, $SE = .04$, $t(1743) = 2.16$, $p = .03$. Linear hypothesis testing showed that both trustworthiness, $F(1, 1743) = 11.38$, $p < .001$, and competence, $F(1, 1743) = 25.53$, $p < .001$, were better predictors than dominance. However, trustworthiness and dominance were not significantly different from each other, $F(1, 1743) = .57$, $p = .45$. For Sample 2 group membership classification, all three traits from Sample 1 were significant predictors: trustworthiness, $\beta = .19$, $SE = .05$, $t(1743) = 8.71$, $p < .001$; competence, $\beta = .46$, $SE = .04$, $t(1743) = 20.58$, $p < .001$; and dominance, $\beta = .19$, $SE = .03$, $t(1743) = 8.43$, $p < .001$. Linear hypothesis testing

showed that competence was the best predictor of group membership [against trustworthiness $F(1, 1743) = 37.57$, $p < .001$; against dominance $F(1, 1743) = 153.75$, $p < .001$)], followed by trustworthiness [against dominance $F(1, 1743) = 8.77$, $p = .003$] and dominance.

Together these results show that the facial regions associated with predicting trustworthiness and competence uniquely relate to facial regions associated with classifying ingroup and outgroup faces, but competence is a better predictor of group membership, indicating that facial stereotypes of group membership consist of facial stereotypes of multiple traits, with the strongest contribution from competence.

GENERAL DISCUSSION

In two studies, we examined the mechanisms underlying mental representations of novel ingroup and outgroup faces. Specifically, we tested whether facial stereotypes of group membership (i.e., facial features that most distinguish ingroup and outgroup faces) consist of visual cues signaling trustworthiness, dominance, and competence. Study 1 suggested that facial stereotypes of group membership indeed resemble facial stereotypes of trustworthiness. Study 2 provided a more complex picture: Facial stereotypes of group membership consist of multiple traits, with the strongest contribution coming from competence. These findings suggest that people can visually imagine the appearance of fellow group members even when these groups are completely novel and not defined by physical appearance. They can do so by relying on attributes that are related to distinct physical characteristics.

Researchers have used the reverse correlation method to understand how people mentally represent social categories and person identities based on physical attributes (Dotsch et al., 2008; Imhoff et al., 2011; Mangini & Biederman, 2004; Young et al., 2014). Other research suggests that people can also imagine the appearance of novel group members even if they have never encountered them before (Hong & Ratner, 2021; Hutchings et al., 2021; Ratner et al., 2014). Although this past work seems to challenge the idea that visual representations require existing memory traces to produce them, our work shows that the visualization of novel group members may be constructed from memory traces of physical attributes that are consistently associated with certain trait impressions (i.e., facial stereotypes). Not surprisingly, facial stereotypes of trustworthiness closely resembled facial stereotypes of group membership. This finding is consistent with previous research suggesting that trust is an important component of intergroup perception (Brewer & Silver, 1978) and that facial trustworthiness predicts whether a novel target will be accepted into the ingroup (Tracy et al., 2020). Our work also shows that facial stereotypes of group membership are determined by multiple traits, with the strongest contribution from competence. This finding indicates that competence, which relates to someone's perceived ability to enact good or bad intentions (Fiske et al., 2007), is an important cue that people rely on to guide visualization of novel ingroup and outgroup members. It is interesting that dominance, which

also similarly signals someone's perceived ability to enact good or bad intentions (Oosterhof & Todorov, 2008), was not consistently associated with facial stereotypes of group membership. Previous research shows that dominance, compared to competence, resembles negative emotion and is perceived as a threat (Said et al., 2009). Perhaps the distinction between novel ingroups and outgroups was driven more by positivity toward the ingroup concept than negativity toward the outgroup concept (Brewer, 1999), and thus trustworthiness and competence, which signal prosociality of the ingroup (but not the outgroup), contributed more to facial stereotypes of group membership.

Although we did not formally examine the meaning of clusters that are important for different facial stereotypes, a visual inspection of the figures (see Figures 2 and 4) clearly shows that important variables are clustered around the eyes, brows, and mouth, corroborating previous research showing that variations in these facial features are associated with trait impressions (e.g., Oosterhof & Todorov, 2008; Zebrowitz et al., 2003). Furthermore, our findings that facial stereotypes of different traits contribute to the visual representation of ingroup and outgroup faces provide a window into how they might indirectly influence face perception. For example, people may focus on the eyes of a face when judging the leadership quality of an ingroup member, whereas they may focus on the brows or lips of a face when judging an outgroup member, which in turn may lead to divergent perception even if they may have similar overall facial features (e.g., high dominance = competence for the ingroup vs. high dominance ≠ competence for the outgroup). Relatedly, recent studies have shown that judgments of facial trustworthiness and facial dominance are more similar for ingroups relative to outgroups (Hong & Freeman, 2023), providing evidence for such a possibility that people may spontaneously look for different facial features when judging the same trait of different groups. It is important to note, however, that it remains unclear whether people are aware of specific strategies and mechanisms for visualizing and making judgments of faces of different social categories. Future research can examine the conscious accessibility of facial stereotypes, which will pave the way for further research on ways to mitigate any potential harmful effects of relying on face characteristics to make dispositional and mental state inferences (Hong et al., 2023). Overall, our findings of biased representation of even novel ingroup and outgroup members have implications for many important social outcomes, including leadership attainment (Todorov et al., 2005) and criminal sentencing (Wilson & Rule, 2015). This is especially true because previous studies have shown that biased representations as measured by reverse correlation methods mediate behavioral outcomes (Lloyd et al., 2020; Ratner et al., 2014).

It is important to note that our results pertain to people's abstract representation of ingroup and outgroup faces. Although our work with two different group membership manipulations suggests that facial stereotypes of competence have the most influence on people's default facial stereotypes of group membership, this effect may be limited to specific instantiations of the minimal group paradigm, and future work will be necessary to examine contextual effects. Even within our own work, the correspondence between facial stereotypes of competence and facial

stereotypes of group membership was more pronounced in Sample 2 than Sample 1. Factors such as whether groups are viewed as having an adversarial relationship to each other or not could influence how much facial stereotypes of ingroups and outgroups resemble facial stereotypes of competence, trustworthiness, and dominance. This would be consistent with work showing that people prefer faces with dominant facial features during intergroup conflict (Hehman et al., 2015). It is further the case that as we develop more elaborated knowledge structures about the physical characteristics associated with specific ingroups and outgroups, as is the case with representations that people have about various racial and ethnic groups, facial stereotypes of group membership might shift away from a strongest reliance on competence facial stereotypes.

Beyond our specific research focus, we introduced methods that can benefit not only other reverse correlation research but also face perception research in general. We used machine learning along with feature selection to parse mechanisms of facial stereotypes of group membership. Our use of OpenFace to extract faces and align facial features across different images makes our analytic approaches easily applicable to other face processing research. For example, analyses of real face images will allow researchers to examine how trait impressions are associated with different facial features across different groups. One limitation of our methods is that we remain agnostic to the meaning of the clusters of variable importance that are associated with trait impressions/group memberships. Our methods were inspired by multivariate pattern analysis that is widely used in neuroimaging studies (e.g., Haxby, 2012). Unlike neuroimaging studies, we lack a platform that synthesizes face image data to help us identify the meaning of clusters (e.g., Neurosynth; Yarkoni et al., 2011). Future research should aim to formally test and identify different locations within a face image as containing meaningful facial features.

Overall, our findings show that competence, rather than trustworthiness and dominance, most resemble facial stereotypes of group membership, providing a nuanced view of ingroup favoritism in face processing that goes above and beyond simple ingroup favoritism. In addition, our work introduces novel analytic methods to analyze face image data in relation to social judgments, allowing researchers to parse mechanisms of social category facial stereotypes.

REFERENCES

Abbas, Z.-A., & Duchaine, B. (2008). The role of holistic processing in judgments of facial attractiveness. *Perception, 37*(8), 1187–1196. https://doi.org/10.1068/p5984

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications.* Technical Report. CMU-CS-16-118. Carnegie Mellon University School of Computer Science.

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*(2), 491–503. https://doi.org/10.1037/a0014201

Antonakis, J., & Eubanks, D. L. (2017). Looking leadership in the face. *Current Directions in Psychological Science, 26*(3), 270–275. https://doi.org/10.1177/0963721417705888

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B., & Silver, M. (1978). Ingroup bias as a function of task characteristics. *European Journal of Social Psychology, 8*(3), 393–400. https://doi.org/10.1002/ejsp.2420080312

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*(10), 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x

Chua, K.-W., & Freeman, J. B. (2021). Facial stereotype bias is mitigated by training. *Social Psychological and Personality Science, 12*(7), 1335–1344. https://doi.org/10.1177/1948550620972550

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. *British Journal of Social Psychology*, *48*(Pt 1), 1–33. https://doi.org/10.1348/014466608X314935

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), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems, 28*, 779–784.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies, 25*(8), 2455–2484. https://doi.org/10.1093/rfs/hhs071

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

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*(6), 878–902.

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science, 20*(4), 419–422. https://doi.org/10.1111/j.1467-9280.2009.02312.x

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage, 62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2015). Picking teams: When dominant facial structure is preferred. *Journal of Experimental Social Psychology, 59*, 51–59. https://doi.org/10.1016/j.jesp.2015.03.007

Hong, Y., Chua, K-W., & Freeman, J. B. (2023). Reducing facial stereotype bias in consequential social judgments: Intervention success with White male faces. *Psychological Science*. Advance online publication. https://doi.org/10.1177/09567976231215238

Hong, Y., & Freeman, J. B. (2023). Shifts in facial impression structures across group boundaries. *Social Psychological and Personality Science.* Advance online publication. https://doi.org/10.1177/19485506231193180

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*(3), 576–600. https://doi.org/10.1037/pspa0000255

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science, 33*(6), 1173–1181. https://doi.org/10.1111/j.1551-6709.2009.01048.x

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. https://doi.org/10.1016/j.cognition.2021.104808

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, *22*(12), 1583–1590. https://doi.org/10.1177/0956797611419675

Kleisner, K., Priplatova, L., Frost, P., & Flegr, J. (2013). Trustworthy-looking face meets brown eyes. *PLoS One, 8*(1), e53285. https://doi.org/10.1371/journal.pone.0053285

Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., & Hugenberg, K. (2020). Good cop, bad cop: Race-based differences in mental representations of police. *Personality and Social Psychology Bulletin, 46*(8), 1205–1218. https://doi.org/10.1177/0146167219898562

Mangini, M., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226. https://doi.org/10.1016/j.cogsci.2003.11.004

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924–940. https://doi.org/10.1037/0278-7393.28.5.924

Oh, D., Walker, M., & Freeman, J. B. (2021). Person knowledge shapes face identity perception. *Cognition, 217*, 104889. https://doi.org/10.1016/j.cognition.2021.104889

Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of Machine Learning Research, 11*, 1833–1863.

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. https://doi.org/10.1002/ejsp.2569

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*(11), 566-570. https://doi.org/10.1016/j.tics.2014.09.007

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. https://doi.org/10.1073/pnas.0805664105

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. https://doi.org/10.1037/a0036498

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260–264. https://doi.org/10.1037/a0014681

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. https://doi.org/10.1002/ejsp.2420010202

Tanis, M., & Postmes, T. (2005). Short Communication: A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology, 35*(3), 413–424. https://doi.org/10.1002/ejsp.256

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Tracy, R. E., Wilson, J. P., Slepian, M. L., & Young, S. G. (2020). Facial trustworthiness predicts ingroup inclusion decisions. *Journal of Experimental Social Psychology, 91*, 104047. https://doi.org/10.1016/j.jesp.2020.104047

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of

human functional neuroimaging data. *Nature Methods, 8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias the mental representation of a presidential candidate's face. *Psychological Science, 25*(2), 503–510. https://doi.org/10.1177/0956797613510717

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review, 7*(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01