This paper was accepted for publication in ***Cognition***. This is a non-final and non-copy-edited version of the paper.

Person knowledge shapes face identity perception

DongWon Oh

New York University

Mirella Walker

University of Basel

Jonathan B. Freeman

New York University

Corresponding author:
DongWon Oh
6 Washington Place
New York, NY 10003
orcid.org/0000-0002-2105-3756
Email: dongwon.oh@nyu.edu

Abstract

Recognition of others' identity through facial features is essential in life. Using both correlational and experimental approaches, we examined how person knowledge biases the perception of others' facial identity. When a participant believed any two individuals were more similar in personality, their faces were perceived to be correspondingly more similar (assessed via mousetracking, Study 1). Further, participants' facial representations of target individuals that were believed to have a more similar personality were found to have a greater physical resemblance (assessed via reverse-correlation, Studies 2 and 3). Finally, when participants learned about novel individuals who had a more similar personality, their faces were visually represented more similarly (Study 4). Together, the findings show that the perception of facial identity is driven not only by facial features but also the person knowledge we have learned about others, biasing it toward alternate identities despite the fact that those identities lack any physical resemblance.

Abstract word count: 149

Keywords: person perception; face processing; semantic memory; mouse tracking; reverse correlation

Person knowledge shapes face identity perception

Recognition of other individuals' faces is essential in life, and people have remarkable face recognition ability (Rossion, 2018; Sunday & Gauthier, 2018; Young & Burton, 2018). Despite holding numerous different individuals' faces in memory, a healthy adult can almost perfectly recognize familiar others within a few seconds (Bruce, 1979, 1983). Much attention has been paid to the cognitive and neural mechanisms underlying this ability (Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Folstein, Palmeri, Van Gulick, & Gauthier, 2015; Haxby, Hoffman, & Gobbini, 2000; Maurer, Le Grand, & Mondloch, 2002). Historically, the majority of facial recognition research has focused on bottom-up perceptual mechanisms, characterizing the visual processes that permit recognition of familiar faces (Johnston & Edmonds, 2009).

However, researchers have also explored how person knowledge may play a role in recognizing facial identity (Bruce, 1983; Bruce & Valentine, 1986; Gordon & Tanaka, 2011; Young, Flude, Hellawell, & Ellis, 1994; Young, Hellawell, & De Haan, 1988). For instance, sequential priming studies have been used to investigate conceptual influences, where faces are presented following related (e.g., Prince Charles and Princess Diana) vs. unrelated faces (e.g., Prince Charles and Hillary Clinton) (Bruce, 1983; Bruce & Valentine, 1986). In these studies, participants are faster to recognize the second face when preceded by a face sharing conceptual overlap, suggesting that person knowledge has the ability to facilitate successful face recognition (Bruce & Young, 1986; Burton et al., 1990). Other studies have demonstrated the faciliatory effects of context on face recognition, such as cases where the scene in which we encounter a person is conceptually related to our stored knowledge about that person (Gruppuso, Lindsay, & Masson, 2007; Mandler, 1980; Winograd & Riversbulkeley, 1977). Such research shows that prior knowledge activated by a prime or contextual cue can speed up the process of recognizing a face, and highlights the role of semantic processes in recognizing familiar faces. These

findings are consistent with connectionist models of face recognition (Bruce & Young, 1986; Burton, Bruce, & Hancock, 1999; Burton et al., 1990) and current models of person perception (Freeman & Ambady, 2011; Freeman, Stolier, & Brooks, 2020), as these models propose an interactive role for semantic (or social-conceptual) representations in recognizing facial identity. Specifically, according these models, after presented with a face, the processing of visual features begins activating identity representations (e.g., Hillary Clinton), and these in turn begin activating social-conceptual representations, such as personality traits (e.g., bold, diligent, competent). Dozens of studies have shown that social-conceptual representations in the form of personality-trait attributes can be automatically activated in response to others' faces (e.g., Kidder, White, Hinojos, Sandoval, & Crites Jr, 2018; Macrae & Martin, 2007). With these social-conceptual representations now activated, such models argue that they then provide feedback to earlier representations in the system, facilitating the activation of associated identity representations (e.g., bold → Hillary Clinton) and in turn the perceptual representation of Clinton's face. Thus, from this perspective, both visual processing and social-conceptual associations can shape the real-time evolution of a face's identity representation, and this allows prior person knowledge and context to adaptively guide face recognition.

One intriguing possibility that arises from these interactive models is the potential biasing effect that may occur when a perceiver incidentally associates two different identities with similar person knowledge. For instance, if a perceiver happens to associate both Hillary Clinton and Elizabeth Warren with similar personality traits (e.g., bold, diligent, competent), when such social-conceptual representations are activated during the processing of Clinton's face, they would feed activation back to not only the Clinton representation but also Warren's and all other associated identity representations. Thus, during the process of recognizing an individual's face, person knowledge that is incidentally shared with an alternate identity could, in theory, activate that alternate identity. This process could bias

the face's perceptual representation more in line with that alternate identity. In other words, the conceptual similarity between Clinton and Warren in the mind of the perceiver could cause Clinton's and Warren's faces to be perceived more similarly as well. Although such biases in perceiving facial identity are a theoretical prediction of current person perception models (Freeman & Ambady, 2011; Freeman & Johnson, 2016) and generally consistent with the premises of longstanding models of face recognition (Bruce & Young, 1986; Burton et al., 1999; Burton et al., 1990), to our knowledge they have never been empirically demonstrated.

In the present research, we test whether person knowledge has the power to shape the perception of a face's identity, biasing it toward alternate identities who are believed to be conceptually similar (in terms of stored person knowledge) yet lack any actual visual resemblance. To do so, we take a representational similarity analysis approach (RSA) (Kriegeskorte, Mur, & Bandettini, 2008). RSA allows a comprehensive assessment of the degree of correspondence across different levels of analysis. Specifically, we assess how the similarity structure of other individuals' identities map onto one another across conceptual, perceptual, and visual levels. For our purposes in the current work, we use "conceptual" to refer to stored conceptual associations about an individual's personality, "perceptual" to refer to how faces are subjectively perceived (based on participants' responses), and "visual" to refer to objective visual properties measured in face images themselves.

In three studies (Studies 1–3), we measured conceptual similarity for all pairwise combinations of different identities based on participants' person knowledge, as well as how similarly those pairwise combinations were perceived (mousetracking study) or visually represented (reverse correlation studies). We then tested whether the extent to which any two identities are deemed conceptually more similar predicts the extent to which faces belonging to those identities are perceived or visually represented more similarly, even when accounting for any intrinsic physical similarity between the two facial

identities. In a final study (Study 4), we directly manipulated conceptual similarity between pairs of identities and measured its corresponding effect on perceptual representations of faces, thereby implicating a causal role of person knowledge in identity perception.

Study 1

We first tested whether conceptual similarity (in the form of personality traits) between familiar individuals predicts perceptual similarity between those individuals. For each participant, perceptual similarity was measured for each pairwise combination of 16 widely known individuals (e.g., celebrities, politicians) using a computer mousetracking technique. Mousetracking is a measure of how multiple perceptual categories activate and resolve over hundreds of milliseconds, allowing a measure of the early processing of faces before a perceptual decision is finalized (Freeman, 2018; Freeman & Ambady, 2010; Freeman, Stolier, Brooks, & Stillerman, 2018). During two-choice mousetracking trials (e.g., a choice between face images of Justin Bieber and Vladimir Putin when presented with Putin's face, **Figure 1**), maximum deviation (MD) in a participant's hand trajectory towards an alternate category response (on the opposite side of the screen) provides an indirect measure of the degree to which the incorrect category was simultaneously "co-activated" during perception despite not being selected. Synchronized mousetracking and neuroimaging has linked such trajectory-deviation effects with neural markers of co-activated categories in brain regions involved in perceptual processing (Brooks, Chikazoe, Sadato, & Freeman, 2019; Stolier & Freeman, 2017). If person knowledge overlaps between two identities, we hypothesize, participants' perceptions will be biased toward the other identity when processing either face. Consequently, their hand trajectories will deviate toward the alternate response for that other identity.

In the mousetracking task, participants were asked to match the identity of a face to the corresponding face as quickly and accurately as possible. As shown in **Figure 1**, the MD from the

straight line connecting the start point to the correct response (correct target face) provides an index of the perceptual similarity between any two given faces (i.e., the extent to which a given target's face was perceived more similarly to the alternate identity). Such biasing effects of person-knowledge overlap should hold above and beyond any potential bottom-up physical resemblance between the two individuals' faces.

Method

Participants. All participants in Study 1 and remaining studies provided informed consent in a manner approved by the University Committee on Activities Involving Human Subjects at New York University. Given no clear precedent for calculating sample size, we aimed to collect a sample of n=200. Two hundred and six individuals living in the US participated via MTurk. Thirteen subjects were excluded for not following instructions, resulting in a final sample of 193 (56.50% male, 42.50% female, 1.04% declined to report; M_{age}=34.60 years, SD_{age}=9.37 years; 67.90% White, 16.60% Hispanic, 6.22% Black, 4.15% Asian, 15.18% other).

To consider the difference in familiarity for each identity pair in our models as a potential confound (see *Familiarity control* in *Procedure* below), we recruited an independent group of raters (n=51) to evaluate their familiarity with each of the 16 target individuals. Eleven participants were excluded for not following instructions, resulting in 40 final raters (65.00% male, 35.00% female; M_{age}=36.80 years, SD_{age}=8.03 years; 2.50% Asian, 10.00%, Black, 27.50% Hispanic, 2.50% Native American, 57.50% White).

To assess the robustness of our findings, in one analysis we reanalyzed our data using conceptual similarity measures derived from an independent group of 499 individuals who participated in a previously published study (Stolier, Hehman, & Freeman, 2020) (35.1% male, 61.7% female, 2.40% other, 0.80% declined to report; M_{age}=30.70 years, SD_{age}=7.15 years; 7.62% Asian, 8.02% Black, 7.41%

Hispanic, 0.80% Native American, 0.60% Pacific Islander, 68.74% White, 4.61% other, 2.20% declined to report).

Stimuli. In the mousetracking task, we presented the 16 target famous individuals' photos, obtained from public-domain websites. In the photos, the individuals were directly oriented and posing natural, relaxed facial expressions (i.e., mildly smiling or resting). No individuals in the photos were wearing glasses, heavy makeup, or had facial tattoos. Facial images were greyscaled and matched in terms of approximate face height and vertical position of the eyes. Extrafacial information (e.g., upper-body clothing and most of the hair) were cropped out. To avoid confounds and reduce the impact of gender and racial stereotypes (Freeman & Ambady, 2011; Macrae & Bodenhausen, 2000), we selected famous individuals with identical gender and race, all male and White: Justin Bieber, George W. Bush, Bill Clinton, Jimmy Fallon, Ryan Gosling, Ashton Kutcher, Matthew McConaughey, Bill Murray, Bill Nye, Vladimir Putin, Keanu Reeves, Robin Thicke, Justin Timberlake, John Travolta, and Mark Wahlberg.

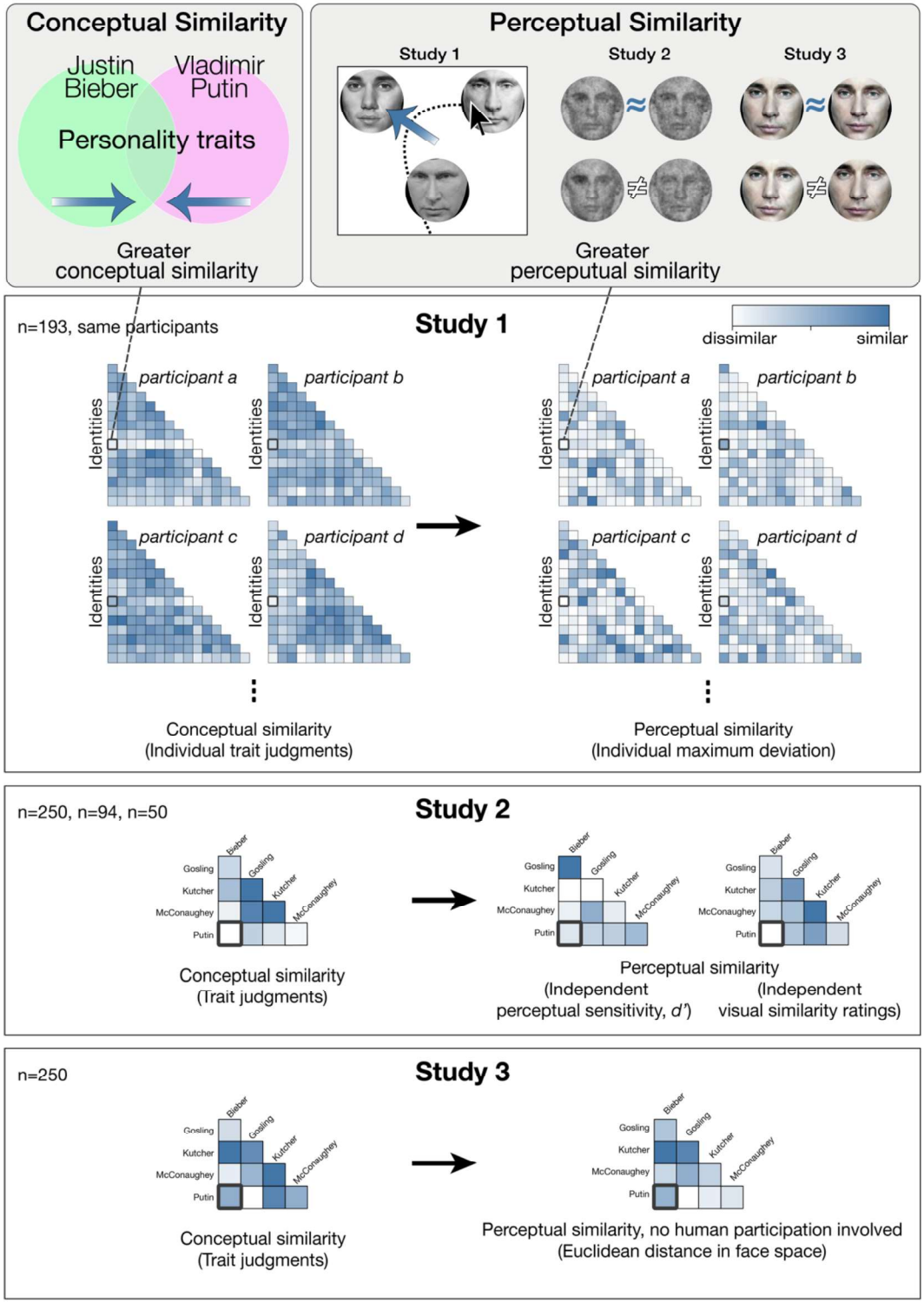

Figure 1. Conceptual and perceptual similarity matrices for Studies 1–3 and the analytic approach. In each study, conceptual similarity and perceptual similarity were assessed for every pair of targets (e.g., Justin Bieber and Vladimir Putin), vectorized, and submitted to multilevel regressions predicting perceptual similarity from conceptual similarity values. The overall hypothesis was that a greater conceptual similarity between any two identities (e.g., Bieber–Putin) would correspond to a

greater bias to perceive the two individuals' faces more similarly, measured by a greater simultaneous attraction to select both targets during mousetracking (Study 1) or a greater resemblance in perceptual representations of the two faces estimated using reverse correlation (Studies 2 and 3). In Study 1, two indices of conceptual similarity – explicit similarity ratings and similarity in target personality trait judgments (the latter is shown in the figure) – predicted the extent to which a given target's face was perceived more similarly to the alternate identity (indexed by mouse-trajectory maximum deviation in the identity match task). Each participant's own conceptual similarity predicted mouse-trajectory maximum deviation in the facial identity match task (two sample participants' similarity matrices are shown in the figure). Conceptual and perceptual identity-pair similarity matrices are displayed for four representative participants. In Study 2, each participant's conceptual similarity of the two targets predicted the extent to which reverse-correlated images were perceptually more similar, assessed using two approaches: the extent to which independent raters could discriminate between the two reverse-correlated images (d') or a separate group of raters' visual similarity ratings of the two images. In Study 3, each participant's conceptual similarity of the two targets predicted the extent to which reverse-correlated images were perceptually more similar, assessed by calculating the distance between the two reverse-correlated faces in terms of their objective facial features. Separate groups of participants participated in Studies 1, 2, and 3. Note that in the matrices for Studies 2 and 3 (reverse-correlation studies), each cell included a different set of participants. Due to the length of the task used, each subject was randomly assigned to be asked about a single identity pair (e.g., Bieber–Putin). Face images used in the description of Study 1 are copyright-free images presented for illustrative purposes, not the actual stimuli.

Response options on each mousetracking trial were face stimuli (i.e., a photo of the foil identity vs. a photo of the correct identity but different from the stimulus photo). For the correct options, we prepared a second facial photo for each of the 16 targets, obtained from public-domain websites. Faces were directly oriented and depicted natural, relaxed facial expressions and contained no glasses, heavy makeup, or facial tattoos. Images were greyscaled, matched in terms of approximate face height and vertical position of the eyes, and extra-facial information was cropped out.

Procedure. Mousetracking data were collected on a JavaScript implementation of MouseTracker software, implementing a standard two-choice design. The 16 face images were each paired with one another, and each pairwise combination was presented twice, resulting in 240 total trials. On each trial, the response options were two facial images (e.g., Justin Bieber's face and Vladimir Putin's face), one of which matched the identity of the face stimulus. Each pair was presented twice because the correct answer alternated between the two identities. On each of 240 trials, participants clicked on a 'Start'

button at the bottom-center of the screen to reveal a face stimulus, which stayed on the screen until they chose one of two response options located in the left and right top corners by moving the mouse cursor and clicking. Participants were encouraged to respond as quickly and accurately as possible. Participants received a warning message to respond more quickly on trials whose duration exceeded 2,500 ms. To ensure trajectories were on-line with the decision process, participants were instructed to begin moving the mouse cursor within 1,500 ms after starting the trial. If they did not begin moving within this time, a message appeared once the trial finished encouraging them to begin moving as they make up their mind. The order of trials and the position of the correct options (left/right) were randomized. At the end of the study, participants were asked to report any identities with whom they were unfamiliar. We later excluded from analysis any trials involving an unrecognized identity (see *Mousetracking data preprocessing* below).

To quantify the conceptual similarity among target individuals, we aimed to provide converging evidence using a two-pronged approach, using both explicit similarity ratings and a measure of personality-trait overlap. First, after completing the mousetracking task, participants rated the pairwise similarity among the targets.[1] Specifically, participants were asked to rate the pairwise similarity among the 16 targets on a scale of 1–7 ("How similar are [person 1] and [person 2]? 1 Not at All Similar–7 Extremely Similar"). On each trial, at the top of the screen participants were instructed: "We are interested in personality impressions of different individuals. In this task, we ask about personality impressions of different well-known individuals, such as politicians and celebrities. While you may not know these individuals directly, we ask you to report how similar these two individuals are to the best of your knowledge and ability. Importantly, go with your gut feeling. We all hold snap personality impressions of others constantly, so feel free to report what you think about these two people." The

[1] Due to a technical error, pairwise similarity ratings of 44 participants were not collected. Results from the remaining participants' data (n=149, 77.20% of all participants) are reported.

instructions ensured that participants would provide a conceptual similarity rating based on similarity in personality impressions (rather than, for example, appearance or occupation).

Second, participants rated specific personality traits of targets. Participants were asked to rate all 16 targets on each of 15 different traits on a 7-point scale ("How [trait] is [person]? 1 Not at all–7 Extremely"): adventurous, angry, anxious, assertive, cautious, cheerful, cooperative, depressed, dutiful, emotional, friendly, intellectual, self-disciplined, sympathetic, trustworthy. We chose these traits to encompass a wide range of individual characteristics that people tend to spontaneously consider when they evaluate faces (Oosterhof & Todorov, 2008; Sutherland et al., 2013) or others' personality (Goldberg, 1999; Stolier et al., 2020; Wiggins, 1979; Wiggins & Pincus, 1992). These 15 traits were chosen from 30 traits ("facets") that compose the Big Five personality factors (3 traits from each factor) (Costa & McCrae, 1992). For the Agreeableness factor, these included cooperation, sympathy, trust; for the Conscientiousness factor, these included dutifulness, self-discipline, cautiousness; for the Extraversion factor, friendliness, assertiveness, cheerfulness; for the Neuroticism factor: anxiety, anger, depression; and for the Openness to Experience factor: emotionality, adventurousness, intellect. These 15 representative facets of the total 30 were previously found to be able to explain various domains of social cognition, including evaluation of both familiar and unfamiliar others (Stolier et al., 2020). Participants were instructed: "While you may not know these individuals directly, we ask you to report how [trait] each person is to the best of your knowledge and ability. Importantly, go with your gut feeling. We all hold snap personality impressions of others constantly, so feel free to report what you think about the person."

For each participant and for each pair of target identities, we computed the euclidean distance between the two 15-item trait-judgment vectors. Thus, higher values indicate that, for the 15 personality traits assessed, two targets' personalities are more dissimilar, and lower values indicate that they were

more similar, conceptually (i.e., in a participant's mind). This procedure permitted the ability to calculate two complementary indices of conceptual similarity for each target identity pair separately for each subject. When these two conceptual similarity measures were acquired, targets' names were presented, absent any faces or other person information. As expected, the two similarity measures of person knowledge (i.e., explicit similarity ratings and personality-trait distance) were positively correlated in our participants (mean r=.28). A one-sample t-test comparing Fisher-Z transformed r coefficients to zero revealed a significant correlation between the two measures (mean Fisher z=0.31, $t(144)=13.59$, 95% CI [0.27,0.36], $p<.001$). While the positive correlation was strongly significant across the participants, indicating that these measures are related to one another, the modest correlation size (r=.28) is consistent with the notion that these are not fully overlapping measures but instead provide complementary information about target individuals' conceptual similarity: one a more global measure of conceptual similarity, and the other a specific trait-distance measure using 15 representative traits.

Mousetracking data preprocessing. Following standard procedures (Freeman & Ambady, 2010), trajectories were normalized into 100 time bins using linear interpolation and rescaled to a standard 2 × 1.5 unit coordinate space. MD of each mouse trajectory towards the incorrect response option on the opposite side of the screen was calculated as the maximum perpendicular deviation relative to an idealized response trajectory (a straight line) drawn between the observed trajectory's start and end point. Trials with incorrect responses, a response time exceeding 2,500 ms, or entailing a face that was reported as unfamiliar to each participant, were excluded from analysis (9.79% of all trials).

Pixel-based visual controls. To account for the potential contribution of bottom-up overlap in physical features between images in any given pair of targets (e.g., physical similarity between Bieber's and Putin's face images used in the mousetracking task), we included visual similarity measures as

covariates in our regression models. Using this extensive set of diverse visual models as covariates ensures that bottom-up visual similarity does not confound conceptual similarity. We calculated two similarity measures derived from pixel values of target facial images. For the first of pixel-based measures, for each pair, we calculated the euclidean distance of the two face stimuli's pixel intensity maps. For the second, we calculated for each pair the euclidean distance of silhouette maps (i.e., binarized pixel values of face or no face). See **Supplementary Figure 1** for all pairwise correlations between pixel-based visual-control covariates and other covariates.

Feature-based visual controls. We calculated six face-based visual similarity measures that better capture similarity in facial features between stimuli. For each face image, we estimated (i) the direction of the left and right eyes, (ii) the head location, (iii) the head rotation, although note all faces were all direct-gaze, center-located, and front-view, (iv) 2D facial landmark coordinates, (v) 3D facial landmark coordinates, and (vi) intensities of facial action units (the specific muscle movements corresponding to different facial expressions) (although note faces were all ostensibly neutral-affect). To estimate these featural measures, we used OpenFace 2.2.0, a robust face-detection algorithm (Baltrušaitis, Zadeh, Lim, & Morency, 2018). Each of the six feature-based visual similarity measures was vectorized for each image. We then calculated for each identity pair the euclidean distance of each measure.

Neural-net-based visual controls. For each pair of images, we calculated three similarity measures derived from three computational models. For the first of neural-net-based measures, we used a model of object recognition (HMAX) (Riesenhuber & Poggio, 1999; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007). Specifically, we used HMAX C2-layer activation, which simulates the neural processing in high-level vision, representing orientation- and size-invariant information of an object. We additionally calculated two similarity measures using more modern, deep convolutional

neural networks (DCNN) specialized for face recognition, quantifying the degree to which the faces appeared similar to the DCNNs for each image pair. A DCNN has multiple hidden layers of interconnected nodes in addition to the input and output layers ("deep"). Many hidden layers are connected to only a subfield of the input layer ("convolutional"), mimicking the topography (i.e., the specificity in the receptive fields) in biological sensory systems. We used two state-of-the-art DCNNs that are pretrained using numerous faces: VGG-Face (2.6 million face images of 2,600 identities) (Parkhi, Vedaldi, & Zisserman, 2015) and Google FaceNet (200 million face images of 8 million identities) (Schroff, Kalenichenko, & Philbin, 2015). Both VGG and FaceNet excel at identity recognition and differentiation, even for untrained face images taken from an unusual angle or under unusual lighting. A face image fed into a DCNN is transformed into a vector in the hyperdimensional space, in which relevant features are determined by the neural net. This vector encodes the individuals' facial identity in the image (O'Toole, Castillo, Parde, Hill, & Chellappa, 2018). For each pair of images, we calculated the euclidean distance between the two vectors.

Familiarity control. We removed any trials for which a participant did not recognize an identity in question (as indicated subsequent to the task; see *Procedure* above). However, even among identities familiar to a participant, there may be graded differences in the level of familiarity across identities. In theory, these familiarity differences across identity could affect our dependent variable (MD). Specifically, if a participant is more familiar with Bieber's face or his past behaviors, than Putin's, for example, their mousetracking trials involving Bieber as the correct answer (vs. trials involving Putin as the correct answer) might reasonably be expected to be facilitated, which would lead to smaller MDs in Bieber trials (vs. Putin trials). To control for this potential confound, we asked an independent group of raters (n=40) to report how familiar they were with each of the 16 target individuals in terms of two types of familiarity: general person knowledge ("what he is like and what he did in the past") and facial

appearance ("what his face looks like") (blocked by familiarity type, 16 questions each). Both types of familiarity were reported on a scale of "Not at All Familiar 1–Extremely Familiar 7" and no face was presented during the task (only name presentation). Both measures showed a high level of inter-rater agreement (general familiarity: ICC=0.78, face familiarity: ICC=0.70), and thus they were averaged separately across the independent raters to provide a target-based familiarity index, which due to the high inter-rater agreement, can be considered sufficiently representative of our primary sample. Expectedly as face familiarity and general familiarity should covary, the two familiarity measures were positively correlated across the 120 targets ($r=.70$, 95% CI [.60,.78], $t(118)=10.71$, $p<.001$). Thus, we averaged the two measures together. For each identity pair presented in the primary task, we calculated the difference in familiarity for the two identities and included this difference score as a covariate in all regression models.

Analytic Approach. We aimed to predict perceptual similarity (MD) from conceptual similarity (explicit similarity rating or similarity in trait judgments, both measured without any face stimuli) while controlling for potential intrinsic visual similarity in the stimuli (overlap in visual properties in the faces measured via pixel-, feature-, and neural-net-based similarity) as well as any potential familiarity difference between the two identities. Perceptual and conceptual similarity values were subject-specific; covariates (visual similarity values and the familiarity difference score) were identical across subjects and specific to each of the 120 pairwise combinations of target identities.

MD for all trials within a subject was rescaled to vary between [0,1] such that 0 corresponded to a subject's largest MD (perceptual similarity) and 1 corresponded to their smallest MD (perceptual dissimilarity). We rescaled MD across trials for each participant to consider individual participants' idiosyncratic response patterns in the mousetracking task (i.e., some participants may have on average

higher MDs than others across trials, and some may lower). Without a standardization across participants, the idiosyncrasy in response patterns can add to statistical noise.

Because two trials corresponded to each target pair (e.g., for Bieber-Putin pair, one in which Bieber was presented and one in which Putin was presented), the two trials' MD values were averaged together, resulting in 120 mean MD values for analysis. These two trials (first and second presentation) were strongly positively correlated ($r=.88$, 95% CI [.84,.92], $t(118)=20.53$, $p<.001$), justifying that they be averaged together to constitute a single perceptual-similarity measure for each identity pair.

Due to the multilevel nature of the data (120 perceptual similarity values nested in each subject), we conducted regression analyses using a multilevel generalized estimating equations (GEE) framework. GEE incorporates nested data while accounting for the inherent dependencies in repeated-measures designs (Liang & Zeger, 1986). We report unstandardized regression coefficients (B) and Wald Z as a measure of effect size.

Results and Discussion

Trials with incorrect responses, a response time exceeding 2,500 ms, or entailing a face that was reported as unfamiliar to a subject, were excluded from analysis (9.79% of all trials). Two GEE multilevel regressions predicting subjects' own perceptual similarity values (MD) from their corresponding conceptual similarity values (indexed separately by explicit similarity ratings and euclidean distance via trait judgments) revealed a strong positive relationship: B=0.010, SE=0.001, 95% CI [0.007, 0.012], Z=7.53, $p<.001$ (explicit pairwise similarity ratings, **Supplementary Figure 2**) and B=0.006, SE=0.001, 95% CI [0.005, 0.008], Z=9.30, $p<.001$ (euclidean distances via trait judgments, **Figure 1**). When we included the models of pixel-, feature-, and neural-net-based visual similarity as well as differences in familiarity across target identities as additional predictors, the relationship between conceptual and perceptual similarity remained significant: B=0.006, SE=0.001, 95% CI [0.003,

0.008], Z=4.63, p<.001 (similarity ratings) and B=0.004, SE=0.001, 95% CI [0.003, 0.005], Z=6.15, p<.001 (euclidean trait distances, **Table 1**).

Another way to test our primary hypothesis is to test the overall model fit when including conceptual measures vs. omitting them, thereby evaluating whether a model that includes conceptual similarity can explain the data better than a model that only includes intrinsic visual similarity and any potential differences in familiarity (covariate-only model). First, we tested whether this covariate-only model significantly explained the perceptual similarity data. If our measures of bottom-up visual similarity actually capture relevant visual information that participants utilize for facial identity perception, then the covariate-only model should significantly capture variance in perceptual similarity. Indeed, the covariate-only model explained the perceptual similarity data significantly better than did an intercept-only model (Wald test for model comparison, $\chi^2(23)=452.00$, $p<.001$). More critically, the full model (including conceptual similarity) explained the perceptual similarity data significantly better than did the covariate-only model ($\chi^2(1)=21.47$, $p<.001$). We obtained the same pattern of results using euclidean trait distance rather than explicit similarity ratings as the conceptual-similarity measure (covariate-only vs. intercept-only model: $\chi^2(23)= 510.12$, $p<.001$; full model vs. covariate-only model: $\chi^2(1)=37.85$, $p<.001$). Together, these results suggest that conceptual person knowledge explained how people perceived facial identities above and beyond any intrinsic facial similarities across individuals as well as potential familiarity differences.

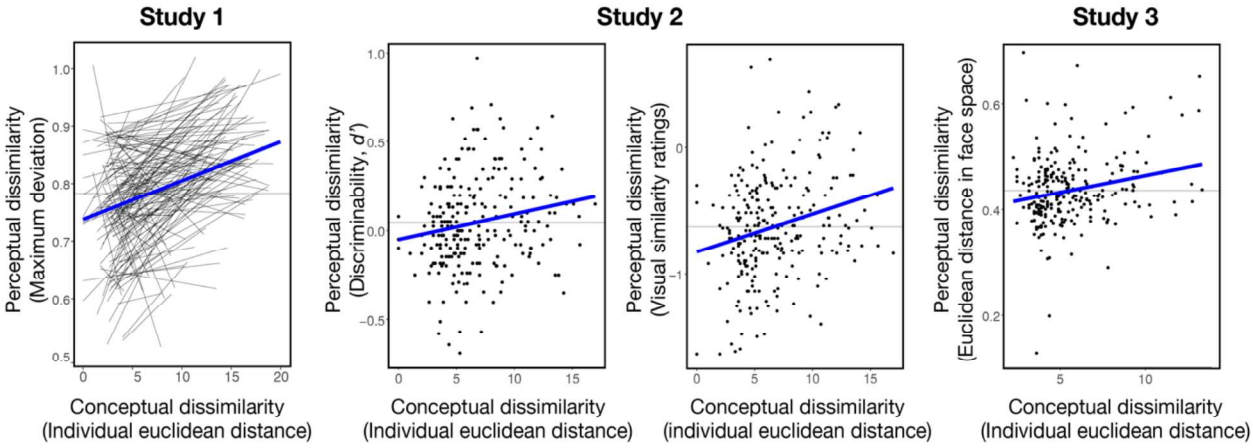

Figure 2. Multilevel regression results in Studies 1–3. In each study, conceptual dissimilarity values predicted perceptual dissimilarity values, while controlling for multiple measures of objective perceptual dissimilarity of the two facial identities. A positive relationship between conceptual and perceptual dissimilarity was observed across studies. For illustrative purposes, the least-squares linear relationship between conceptual and perceptual dissimilarity is plotted for each study. Actual analyses were conducted using GEE multilevel regressions. Blue lines denote the average relationship across all participants and targets. For Study 1, lines denote slopes for individual participants, representing the relationship between conceptual and perceptual similarity collected within the same participants. Dots denote participants in Studies 2–4. Individual lines are not displayed for these studies because conceptual and perceptual similarity was only measured for a single identity pair for each participant.

To further assess the robustness of the primary effect of interest, we reran analyses using a third measure of conceptual similarity available from a previously published dataset (15-trait-judgment euclidean distances among the same target individuals), derived from independent raters (n=499) rather than the participants themselves (data available at https://osf.io/2uzsx) (Stolier et al., 2020). When using these independent raters' conceptual similarity measures (rather than participants' own conceptual similarity measures) and including visual and familiarity covariates as additional predictors, GEE regression again revealed a strong positive relationship: B=0.010, SE=0.001, 95% CI [0.008, 0.012], Z=8.44, p<.001.

To demonstrate that meaningful individual differences in participants' person knowledge about identities predicted corresponding differences in facial identity perception, two additional analyses were conducted. This is because it is possible that the results of the analyses thus far could be explained by

average conceptual and perceptual tendencies across the sample, rather than idiosyncratic variability across participants. In a first stringent analysis, we calculated a group-average conceptual similarity measure by averaging conceptual similarity values across participants and included this as a covariate in the regression models. An effect of subject-specific conceptual similarity that holds above and beyond the group-average values would show that unique differences across participants explain a significant amount of the remaining variance in perceptual similarity. While group-average conceptual similarity was a positive and significant predictor of perceptual similarity (similarity ratings: B=0.031, SE=0.004, 95% CI [0.023, 0.038], Z=7.96, p<.001; euclidean trait distance: B=0.019, SE=0.002, 95% CI [0.014, 0.023], Z=8.86, p<.001), more critically, subject-specific conceptual similarity remained a strong positive predictor (similarity ratings: B=0.003, SE=0.001, 95% CI [0, 0.005], Z=2.06, p=.039; euclidean trait distance: B=0.002, SE=0.001, 95% CI [0.000, 0.003], Z=2.61, p=.009). Thus, while there were certainly average tendencies that existed across the sample, there was unique variability across participants and these individual differences contributed to corresponding differences in facial identity perception.

In a second complementary analysis, we clustered the data by identity-pair rather than by subject in the multilevel regression. Accordingly, an effect of conceptual similarity would show that within a given identity pair (e.g., Bieber–Putin), subjects with stronger conceptual similarity (e.g., deeming Bieber and Putin as having similar personalities) are also those subjects who exhibit stronger perceptual similarity (e.g., perceiving Bieber and Putin's faces as similar). If there were no meaningful variability across subjects, then the effect of conceptual similarity would not yield a significant result. Consistent with prior work (Brooks & Freeman, 2018), all conceptual and perceptual values were first ranked within each subject, thereby removing possible differences in overall magnitude or scale in these variables across subjects (i.e., the possibility that some subjects have higher/lower similarity values

overall or more/less dispersion, across all 120 identity pairs). Ranking vs. non-ranking values had no influence on the pattern of results. Identity pairs' perceptual similarity values were regressed onto their conceptual similarity values with all covariates included, as in the main analysis. The results revealed a positive relationship (similarity ratings: B=0.055, SE=0.010, 95% CI [0.036, 0.074], Z=5.69, p<.001, euclidean trait distance: B=0.055, SE=0.008, 95% CI [0.040, 0.070], Z=7.15, p<.001). Converging with the previous analysis that controlled for group-average conceptual similarity, these results show that reliable individual differences exist in participants' person knowledge which manifests as corresponding differences in facial identity perception, above and beyond any average tendencies across the sample.

Overall, the findings of Study 1 demonstrate that the extent to which a participant believes two targets have a more similar personality predicts how similarly that participant perceives those targets' faces to be, regardless of whatever physical similarity may exist between the two faces. Moreover, the results show that individual differences in a perceiver's own unique person knowledge predicts corresponding differences in facial identity perception.

Table 1. Full Model Statistics in Study 1. Perceptual dissimilarity ([0,1], derived from maximum deviation in the mouse trajectory in a mousetracking task) was predicted from individual participants' own pairwise dissimilarity rating (left) and euclidean distance based on their trait ratings of target identities (right) in two separate GEE models.

Predictor	B	SE	Z	p	Predictor	B	SE	Z	p
Conceptual: similarity rating	0.006	0.001	4.63	<.001	Conceptual: trait distance	0.004	0.001	6.15	<.001
Pixel: intensity Center	<0.001	<0.001	2.40	.016	Pixel: intensity Center	<0.001	<0.001	1.46	.144
Pixel: intensity Option	0.001	<0.001	6.22	<.001	Pixel: intensity Option	0.001	<0.001	7.62	<.001
Pixel: silhouette Center	<0.001	<0.001	4.61	<.001	Pixel: silhouette Center	<0.001	<0.001	4.98	<.001
Pixel: silhouette Option	<0.001	<0.001	0.76	.450	Pixel: silhouette Option	<0.001	<0.001	0.98	.329
Feature: gaze Center	0.235	0.03	7.74	<.001	Feature: gaze Center	0.235	0.027	8.61	<.001
Feature: gaze Option	–0.162	0.034	4.71	<.001	Feature: gaze Option	–0.163	0.029	5.55	<.001
Feature: head location Center	0.001	0.001	0.56	.576	Feature: head location Center	0.001	0.001	0.68	.496
Feature: head location Option	–0.009	0.001	7.26	<.001	Feature: head location Option	–0.008	0.001	7.87	<.001
Feature: head rotation Center	–0.133	0.032	4.16	<.001	Feature: head rotation Center	–0.121	0.028	4.27	<.001

Feature: head rotation Option	–0.030	0.036	0.84	.401	Feature: head rotation Option	–0.034	0.031	1.11	.265
Feature: landmark 2D Center	<0.001	<0.001	5.55	<.001	Feature: landmark 2D Center	<0.001	<0.001	6.66	<.001
Feature: landmark 2D Option	<0.001	<0.001	0.92	.357	Feature: landmark 2D Option	<0.001	<0.001	0.73	.463
Feature: landmark 3D Center	<0.001	<0.001	0.85	.398	Feature: landmark 3D Center	<0.001	<0.001	1.01	.315
Feature: landmark 3D Option	0.001	<0.001	7.27	<.001	Feature: landmark 3D Option	0.001	<0.001	8.11	<.001
Feature: action unit Center	0.001	0.001	1.05	.295	Feature: action unit Center	0.001	0.001	1.02	.309
Feature: action unit Option	0.003	0.002	1.62	.106	Feature: action unit Option	0.003	0.001	2.42	.015
Neural net: HMAX C2 Center	–0.007	0.002	3.12	.002	Neural net: HMAX C2 Center	–0.008	0.002	4.41	<.001
Neural net: HMAX C2 Option	0.013	0.002	5.70	<.001	Neural net: HMAX C2 Option	0.012	0.002	6.22	<.001
Neural net: FaceNet Center	–0.002	0.001	1.50	.135	Neural net: FaceNet Center	–0.001	0.001	0.70	.485
Neural net: FaceNet Option	0.001	0.001	0.82	.413	Neural net: FaceNet Option	0.001	0.001	0.77	.442
Neural net: VGG Center	0.107	0.035	3.06	.002	Neural net: VGG Center	0.111	0.031	3.55	<.001
Neural net: VGG Option	–0.020	0.042	0.49	.628	Neural net: VGG Option	–0.056	0.035	1.61	.108
Familiarity	0.008	0.006	1.34	.182	Familiarity	0.009	0.005	1.75	.081

Note: The predictors of interest are in the first row (similarity rating: individual participants' pairwise dissimilarity rating between identities, trait distance: individual participants' euclidean distance of two identities based on trait ratings). In Study 1, face images were presented as options ('Option') as well as the center reference image ('Center') in a perceptual matching mousetracking task. B=unstandardized regression coefficient, SE=standard error, Z=Wald Z.

Study 2

In Study 2 and the remaining studies, we provide converging evidence for the impact of person knowledge on facial identity perception using a data-driven approach that is less constrained to particular face stimuli. We also provide a stronger test of the hypothesis that the faces belonging to target identities believed to have a more similar personality are actually "seen" more similarly. To do so, we used a reverse correlation technique. Devised first as a tool to identify stimulus features that contribute to a perceptual decision (Ahumada Jr & Lovell, 1971), reverse correlation became a valuable approach in vision science (for review, see Eckstein & Ahumada, 2002). More recently, reverse correlation gained popularity as a tool to visualize facial characteristics that contribute to a decision

about social attributes (for reviews, see Brinkman, Todorov, & Dotsch, 2017; Dotsch & Todorov, 2011). Reverse correlation is able to approximate a perceptual representation of a target category (i.e., a template or prototype of the category in the observer's mind). For our purposes, 'perceptual representation' refers to a visualized mental representation or prototype of facial identity assessed through reverse-correlation techniques. In Study 2, we superimpose random noise patterns over a single base face and ask participants to select which of two noise-altered face images appear to better match a target's face. Averaging the noise patterns reveals an estimate of how a target's face appears in the mind's eye of a subject. We hypothesize that the extent to which a perceiver's knowledge of two identities' personality is more similar will correspond to a greater resemblance in that perceiver's representation of their faces.

Method

Participants. Because reverse correlation tasks require a large number of trials per condition, participants were asked to each complete only one identity-pair condition (e.g., Bieber and Putin). We aimed to recruit 25 subjects for each identity-pair. A total of 252 individuals living in the US participated via MTurk. Two participants were excluded for not following instructions, resulting in a final sample of 250 (54.40% male, 44.80% female, 0.40% other, 0.40% declined to report; M_{age}=33.92 years, SD_{age}=9.70 years; 79.20% White, 10.80% Black, 3.60% Asian, 6.40% other). Once reverse-correlated images were generated from the subjects, two groups of independent raters were recruited to either categorize these images as one facial identity vs. a foil (n=94; 56.38% male, 42.55% female, 1.06% declined to report; M_{age}=33.60 years, SD_{age}=9.19 years; 5.32% Asian, 7.45% Black, 20.21% Hispanic, 1.06% other, 61.70% White, 4.26% declined to report), or rate all pairwise combinations of images on their visual similarity (n=50; 56.00% male, 44.00% female, M_{age}=39.40 years, SD_{age}=14.30

years; 2.00% Asian, 14.00% Hispanic, 82.00% White, 2.00% other). All participants followed the instructions, and no participants were excluded from analysis.

Stimuli. To create the reverse correlation stimuli, we first chose 5 individuals with a range of conceptual similarity (based on the average data from Study 1): Justin Bieber, Ryan Gosling, Ashton Kutcher, Matthew McConaughey, and Vladimir Putin. Using the face stimuli of Study 1, we then created the base face image of each pairwise identity combination by morphing the two faces to a 50/50 blend of each identity using PsychoMorph (Tiddeman, Burt, & Perrett, 2001). Consistent with previous reverse correlation studies (Brinkman et al., 2017), we then applied to each base image a Gaussian blur with 3-pixel radius, removing high spatial frequency information. We then imposed five layers of sinusoid noise patterns varying in spatial scale and their negative versions using the *rcicr* R library (Dotsch, 2015). For the 50/50 base image of each identity pair, we created 400 images comprising of 200 side-by-side face trials. See **Supplementary Figure 3** for details of stimulus preparation and reverse-correlation task procedures.

Procedure. The reverse correlation task followed standard procedures from previous studies (Brinkman et al., 2017; Dotsch & Todorov, 2011). Each participant in the task was assigned one of the 10 pairwise combinations of the 5 target identities (e.g., Bieber and Putin). Following the procedure of prior reverse correlation research involving facial identity (Mangini & Biederman, 2004), participants were instructed: "You will see pictures of [person A] and [person B] that are warped to the same geometry. That is, their eyes, nose, mouth, hairline, chin, etc. are in identical locations. Your task is to discriminate between [person A] and [person B]. The task is difficult because of the warping manipulation and the visual degradation. Go with your intuition, respond to the best of your ability." On each trial in the reverse correlation task, participants were presented with two side-by-side noise-imposed face images (see *Stimuli* above). The task was split into two blocks each of 200 trials.

A participant in the Bieber-Putin condition, for example, was asked to complete 200 trials where they were instructed to choose the face that "appears more like Justin Bieber" and 200 trials where they were instructed to choose the face that "appears more like Vladimir Putin". The trial order and task identity order were randomized. Following the standard preprocessing procedure (Dotsch & Todorov, 2011), we averaged the noise-imposed face selected on each trial for each participants, resulting in reverse-correlated images for each of the two identities each subject classified. Across all subjects, this resulted in a total of 500 reverse-correlated images (2 identities × 10 identity pair conditions × 25 subjects per condition). After completing the reverse correlation task, participants answered an explicit pairwise similarity rating question and completed the personality trait rating task used in Study 1. However, instead of rating the 16 targets as in Study 1, they only rated the two targets in their assigned pair condition.

Image similarity in each participant's two reverse-correlated images was assessed using two complementary methods. One independent group of raters (n=97) categorized the reverse-correlated images by identity in a forced-choice task. The 500 reverse-correlated images were randomly divided into two sets of 250 reverse-correlated images (each derived from 5 identity-pair conditions) and assigned to two different subgroups (n=50 and n=47). On each trial, participants were presented with a reverse-correlated image and two names (e.g., "Justin Bieber" vs. "Vladimir Putin") and instructed to decide whether the face looks like one person or the other by the corresponding identity. Trials were blocked by identity-pair condition. Raters were encouraged to use the full set of options (e.g., both "Bieber" option and "Putin" option).

A separate group of raters (n=50) rated each of the 250 total pairs of reverse-correlated images on similarity. Each rater judged the similarity of all 250 reverse-correlated image pairs. Each reverse-correlated image pair was generated from an initial subject of the reverse correlation task. On each trial,

participants were instructed to rate the two images on how visually similar they were on a 7-point scale ("Please rate how similar the two faces appear. 1 Not at All Similar–7 Extremely Similar"). Raters were encouraged to use the full set of options. The origin of the images or the underlying identities were not mentioned.

Analytic approach. We assessed the relationship between each individual participant's personality knowledge about identities (conceptual similarity) and the appearance of their reverse-correlated images (perceptual similarity). To do so, we conducted GEE regression analyses testing whether the degree of overlap in conceptual knowledge between a given pair of identities (e.g., conceptual similarity between Bieber and Putin) corresponds to a biased perceptual resemblance in the appearance of the two reverse-correlated images (e.g., how similar Bieber's and Putin's reverse-correlated images appeared).

In each regression model, we clustered data by the identity pair to consider the inherent variation in physical facial similarity in each pair. To further control for the potential effect of baseline similarity between the two individuals (e.g., physical facial similarity between Bieber and Putin), as in Study 1, we included the model measures of visual similarity between the two original facial photos used to create the morphed base image for reverse correlation as well as the familiarity difference score between the two identities (see Study 1 Method for details). We included all visual similarity measures used in the previous studies: pixel- (pixel intensity, silhouette), feature- (gaze, head location, head rotation, 2D and 3D landmarks, action units), and neural-net-based dissimilarity measures (HMAX C2, FaceNet, VGG-Face) in euclidean distance. See **Supplementary Figure 4** for all pairwise correlations between all covariates, including the visual-control covariates and familiarity covariate.

Consistent with Study 1, conceptual similarity was measured via two complementary indices: the pairwise similarity rating and the euclidean distance between the two identities' 15-item trait rating

vectors. Perceptual similarity in the reverse-correlated images was assessed via two indices from independent raters who evaluated the reverse-correlated images. For the first similarity index, independent raters categorized the identity of all reverse-correlated images in a two-alternative, forced-choice task. The perceptual similarity for a given identity pair was provided by independent raters' average perceptual discriminability (d') of the two reverse-correlated images on the trials where they were pitted against one another. Specifically, for each of the two reverse-correlated images in an identity pair (e.g., Bieber reverse-correlated image and Putin reverse-correlated image), one identity was arbitrarily defined as signal (e.g., Putin) and the other as noise (e.g., Bieber). For example, a hit would be defined as categorizing a Putin reverse-correlated image as Putin, a miss as categorizing a Putin reverse-correlated image as Bieber, a correct rejection as categorizing a Bieber reverse-correlated image as Bieber, and a false alarm as categorizing a Bieber reverse-correlated image as Putin, thereby giving d' for each identity pair (Tanner & Swets, 1954). The average d' across raters served as each identity pair's index of perceptual similarity. For the second perceptual similarity index, another set of independent raters directly rated the apparent similarity of all pairwise combinations of reverse-correlated images. Mean similarity ratings were recoded such that –3 indicated a maximal perceptual similarity and 3 indicated maximal perceptual dissimilarity.

Results and Discussion

Regardless of the measure used, GEE regressions all revealed that conceptual similarity between a given pair of identities strongly predicted the perceptual similarity of those two identities' reverse-correlated image.[2] Ratings of conceptual similarity predicted independent raters' reduced ability to

[2] The multilevel regression models in Studies 2–4 initially failed to converge due to multicollinearity between a subset of the visual-model covariate predictors. Visual model covariate predictors were excluded one at a time, starting with those with the strongest correlations with other predictors (defined as the biggest sum of the absolute value of Fisher-Z-transformed pairwise correlation coefficients, which was calculated at each iteration), until each model converged. The results reported here describe the models when four (Studies 2 and 3) and eight (Study 4) visual-model covariate predictors were removed following this exclusion rule. Running the models using all other sets of excluded visual model covariate predictors did not

perceptually discriminate (d') between the two reverse-correlated image (B=0.020, SE=0.005, 95% CI [0.009, 0.030], Z=3.69, p<.001) and predicted stronger ratings of visual similarity (B=0.029, SE=0.006, 95% CI [0.016, 0.042], Z=4.48, p<.001, **Supplementary Figure 5**). Euclidean trait similarity also predicted independent raters' d' (B=0.015, SE=0.003, 95% CI [0.008, 0.021], Z=4.19, p<.001) and visual similarity ratings (B=0.014, SE=0.004, 95% CI [0.006, 0.022], Z=3.55, p<.001, **Figures 1–3**).

When we included covariates (visual similarity estimates and familiarity difference) between the original facial photos used to create the noise-imposed images (prior to morphing) – the relationship between conceptual and perceptual similarity remained significant in all regression models: B=0.018, SE=0.006, 95% CI [0.007, 0.029], Z=3.19, p=.001 (conceptual similarity ratings predicting d'), B=0.027, SE=0.007, 95% CI [0.013, 0.041], Z=3.79, p<.001 (conceptual similarity ratings predicting perceptual similarity ratings), B=0.019, SE=0.004, 95% CI [0.011, 0.026], Z=4.98, p<.001 (euclidean trait similarity predicting d'), and B=0.013, SE=0.004, 95% CI [0.005, 0.021], Z=3.12, p=.002 (euclidean trait similarity predicting perceptual similarity ratings, **Tables 2 & 3**).

We tested, as in Study 1, whether a model that includes conceptual similarity vs. omits it better explains the perceptual similarity data. Each full model (including conceptual similarity) indeed explained the data better than the corresponding model that omitted it (conceptual similarity ratings predicting d': $\chi^2(1)=10.16$, p=.001, conceptual similarity ratings predicting perceptual similarity ratings: $\chi^2(1)=14.39$, p<.001, euclidean trait similarity predicting d': $\chi^2(1)=24.81$, p<.001, euclidean trait similarity predicting perceptual similarity ratings: $\chi^2(1)=9.73$, p=.002). Expectedly, the covariate-only model also explained the data better than an intercept-only model (predicting d': $\chi^2(8)=1582.06$, p<.001; predicting perceptual similarity ratings: $\chi^2(8)=78.10$, p<.001), indicating that the bottom-up visual model covariates, along with the familiarity covariate, accurately captured visual information utilized for facial

meaningfully change the results. See **Supplementary Figures 1**, **4**, **7**, & **10** for pairwise correlations between all covariates in studies 1–4, respectively.

identity perception. These results suggest that person knowledge explained perceptions of facial identities above and beyond any potential similarity intrinsic to the targets' faces or differences in familiarity.

As the conceptual and perceptual measures in Study 2 were collected from separate groups of participants, and 'conceptual measure' participants (those who participated in a reverse-correlation task) contributed to only one identity pair, there was no need in Study 2 to conduct additional individual-difference analyses (group-average covariate and identity-pair clustering). Because our analyses already clustered the data by identity pair, these results could only yield a significant result if reliable individual differences in conceptual similarity existed across our reverse-correlation participants, which manifested as corresponding individual differences in perceptual similarity.

In sum, across multiple converging measures of conceptual similarity and measures of perceptual similarity, the findings of Study 2 show that individual differences in the extent to which any two targets are believed to have a more similar personality predicts a greater approximation in how those identities' faces are perceptually represented in a participant's mind (**Figure 3**).

Table 2. Full Model Statistics in Study 2. Perceptual discriminability (d') was predicted from individual participants' pairwise dissimilarity rating (left) and euclidean distance based on their trait ratings of target identities (right) in two separate GEE models.

Predictor	B	SE	Z	p	Predictor	B	SE	Z	p
Conceptual: similarity rating	0.018	0.006	3.19	.001	Conceptual: trait distance	0.019	0.004	4.98	<.001
Pixel: intensity	0.003	0.001	4.28	<.001	Pixel: intensity	0.003	0.001	5.68	<.001
Feature: gaze	-1.173	0.300	3.91	<.001	Feature: gaze	-1.071	0.230	4.67	<.001
Feature: landmark 2D	<0.001	<0.001	0.22	.829	Feature: landmark 2D	<0.001	<0.001	0.63	.531
Feature: landmark 3D	<0.001	<0.001	1.52	.130	Feature: landmark 3D	<0.001	<0.001	1.79	.073
Feature: action unit	–0.035	0.006	5.52	<.001	Feature: action unit	–0.035	0.005	7.14	<.001
Neural net: HMAX C2	–0.040	0.011	3.58	<.001	Neural net: HMAX C2	–0.047	0.008	5.92	<.001
Neural net: FaceNet	<0.001	0.008	0.01	.993	Neural net: FaceNet	0.006	0.006	0.96	.335
Familiarity	–0.148	0.038	3.93	<.001	Familiarity	–0.193	0.026	7.47	<.001

Note: The predictors of interest are in the first row (similarity rating: individual participants' pairwise dissimilarity rating between identities, trait distance: individual participants' euclidean distance of two identities based on trait ratings). B=unstandardized regression coefficient, SE=standardized error, Z=Wald Z.

Table 3. Full Model Statistics in Study 2. Perceptual dissimilarity ([–3,3], derived from pairwise dissimilarity ratings) were predicted from individual participants' pairwise dissimilarity rating (left) and euclidean distance based on their trait ratings of target identities (right) in two separate GEE models.

Predictor	B	SE	Z	p	Predictor	B	SE	Z	p
Conceptual: similarity rating	0.027	0.007	3.79	<.001	Conceptual: trait distance	0.013	0.004	3.12	.002
Pixel: intensity	–0.002	0.004	0.45	.652	Pixel: intensity	–0.002	0.004	0.40	.687
Feature: gaze	1.680	1.703	0.99	.324	Feature: gaze	1.737	1.652	1.05	.293
Feature: landmark 2D	0.001	0.002	0.26	.792	Feature: landmark 2D	0.001	0.002	0.24	.812
Feature: landmark 3D	<0.001	0.001	0.32	.750	Feature: landmark 3D	0.001	0.001	0.38	.705
Feature: action unit	–0.001	0.036	0.04	.970	Feature: action unit	–0.003	0.035	0.08	.934
Neural net: HMAX C2	0.091	0.065	1.40	.160	Neural net: HMAX C2	0.088	0.062	1.42	.156
Neural net: FaceNet	0.005	0.044	0.12	.905	Neural net: FaceNet	0.012	0.042	0.28	.776
Familiarity	0.534	0.213	2.50	.012	Familiarity	0.502	0.208	2.42	.016

Note: The predictors of interest are in the first row (similarity rating: individual participants' pairwise dissimilarity rating between identities, trait distance: individual participants' euclidean distance of two identities based on trait ratings). B=unstandardized regression coefficient, SE=standard error, Z=Wald Z.

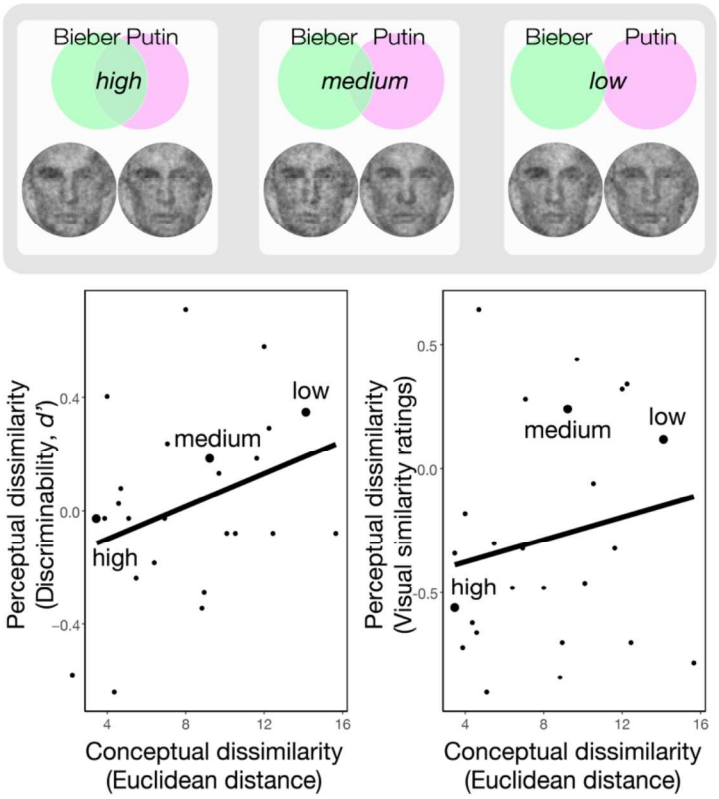

Figure 3. Reverse correlation results in Study 2. To provide an example of the pattern of results, reverse-correlated images for one identity pair are depicted (Bieber and Putin) separately for high, average, and low tertiles of participants' conceptual similarity (i.e., overlapping person knowledge) (top). Participants with higher conceptual similarity between Bieber and Putin produced reverse-correlated images of Bieber and Putin that exhibited a greater perceptual resemblance, as assessed by independent raters' inability to discriminate between the two reverse-correlated images (bottom left) and their stronger ratings of visual similarity (bottom right). Each dot in the scatterplots represents a single reverse correlation task participant, and the solid line denotes the linear fit. The plots are displayed for illustrative purposes. Actual analyses were conducted using GEE multilevel regression.

Study 3

Study 2 showed that when a participant regarded two individuals' personalities as more similar, estimated perceptual representations of their faces were correspondingly more similar. However, Study 2 employed a noise-imposing reverse correlation technique which afforded limited analysis of specific facial features driving identification. In Studies 3 and 4, we corroborate our findings using a different type of reverse correlation technique, in which a face is determined by a large set of shape and color features via a face model (Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009; Walker & Vetter,

2009). In this framework, a face is represented as a vector in a multidimensional face space (Valentine, 1991). Not only does a reverse correlation approach based on such a multidimensional face model generates more realistic faces, but more critically it allows the objective specification of how facial features become increasingly resembling as two different identities' conceptual similarity increases.

Method

Participants. As in Study 2, we aimed to recruit 25 subjects for each identity pair. We employed the same five well-known White male individuals as target identities, again resulting in 10 identity-pair conditions. Two hundred and sixty-four individuals living in the US participated via MTurk. Fourteen participants were excluded for not following instructions, resulting in a final sample of 250 (70.80% male, 28.80% female, 0.40% other, M_{age}=38.80 years, SD_{age}=8.83 years, 47.20% White, 39.20% Black, 11.20% Asian, 2.00% other).

Stimuli. To create the face prototype of each of the five target individuals, we used WebMorph (DeBruine, 2018), an upgraded and online adaptation of PsychoMorph (Tiddeman et al., 2001). We averaged three images of each of the five individuals found on the web, and standardized all five images so the faces were front-facing, using OpenFace 2.2.0 (Baltrušaitis et al., 2018). We then created the base face image of each pairwise identity combination (i.e., the 50/50 blend of each identity for each pair). Once we had these morphs, we transformed them into vectors in the Basel Face Model face space (Paysan et al., 2009). The Basel Face Model employs a multidimensional space in which each dimension represents a change in a featural variation on the face's shape and color based on laser-scanned data of actual human faces. To create random stimuli for the reverse correlation task, we created variations of each morphed face applying previously generated random vectors (Walker & Keller, 2019), generating 100 random face pairs varying on shape and 100 random face pairs varying on color. Two faces in each pair were manipulated in the opposite direction in the face space in relation to the midpoint of the face

space. This step is analogous to the noise-imposing procedure on the base morph faces in Study 2. Resulting random faces that varied on shape or color resembled either of the two identities, some more strongly than the others. As in Study 2, for each 50/50 base face for each identity pair, we created 400 images that were paired to each other, resulting in 200 side-by-side face trials. See **Supplementary Figure 6** for details of stimulus preparation and reverse-correlation task procedures.

Procedure. As in Study 2, the task for each identity entailed 200 trials. After data collection, we averaged the full set of shape and color parameters of the face selected on each trial for each participant, resulting in reverse-correlated face vectors for each of the two identities that each participant classified.

Analytic approach. As in Study 2, we predicted conceptual similarity from perceptual similarity via multilevel regressions with the same set of covariates (see **Supplementary Figure 7** for all pairwise correlations between all covariates). A key difference was that we calculated perceptual similarity as the euclidean distance between the resulting two reverse-correlated vectors in the face space for each participant. This euclidean distance represented how visually similar the two faces were between the two identities at the level of objective facial features for each subject in their mind's eye.

Results and Discussion

GEE regressions tested the relationship between conceptual similarity (assessed separately via similarity rating and euclidean trait distance) and objective perceptual similarity (assessed via euclidean distance between the two reverse-correlated images' full set of facial features in the face space). Indeed, ratings of conceptual similarity between a pair of identities predicted a greater resemblance between the two identities' reverse-correlated images at the level of their objective features (B=0.016, SE=0.003, 95% CI [0.010, 0.023], Z=5.28, p<.001, **Supplementary Figure 8**). Euclidean trait similarity also predicted greater resemblance at the level of the reverse-correlated images' objective features (B=0.006, SE=0.002, 95% CI [0.003, 0.010], Z=3.50, p<.001, **Figures 1 & 2**).

When we included various similarity estimates of the original facial photos used prior to morphing as well as the familiarity difference measure as covariates, the relationship between conceptual similarity of any given pair of identities and objective resemblance of their reverse-correlated images remained significant (conceptual similarity ratings: B=0.017, SE=0.003, 95% CI [0.011, 0.023], Z=5.73, p<.001) (euclidean trait similarity: B=0.006, SE=0.002, 95% CI [0.003, 0.009], Z=3.58, p<.001, **Table 4**).

We tested, as in preceding studies, whether a model that includes conceptual similarity explained the data better than did the same model omitting conceptual similarity. Indeed, the full model explained the data better than a model omitting conceptual similarity, both in conceptual similarity ratings predicting perceptual similarity ($\chi^2(1)=32.84$, p<.001) and euclidean trait similarity predicting perceptual similarity ($\chi^2(1)=12.78$, p<.001). Expectedly, the covariate-only model also explained the data better than an intercept-only model ($\chi^2(8)=88.14$, p<.001), indicating that the visual model covariates, along with the familiarity covariate, were capturing visual information actually utilized for perception. These results suggest that person knowledge explained how people perceived facial identities above and beyond the similarity between individuals' faces and the level of familiarity.

The results converged with Study 2 but here used a reverse correlation technique that not only relies on more realistic face images but permits objective analysis of featural resemblance. When a participant believed a given pair of identities have more overlapping person knowledge, their perceptual representations of the two identities' faces became more featurally resembling.

Table 4. Full Model Statistics in Study 3. Perceptual face-space distances were predicted from individual participants' pairwise dissimilarity rating (left) and euclidean distance based on their trait ratings of target identities (right) in two separate GEE models.

Predictor	B	SE	Z	p	Predictor	B	SE	Z	p
Conceptual: similarity rating	0.017	0.003	5.73	<.001	Conceptual: trait distance	0.006	0.002	3.58	<.001

Pixel: intensity	<0.001	<0.001	2.53	.012	Pixel: intensity	<0.001	<0.001	2.09	.037
Pixel: silhouette	0.001	<0.001	3.13	.002	Pixel: silhouette	<0.001	<0.001	2.50	.012
Feature: gaze	–0.204	0.131	1.56	.119	Feature: gaze	–0.121	0.144	0.84	.400
Feature: head location	–0.014	0.004	3.17	.002	Feature: head location	–0.011	0.005	2.31	.021
Feature: landmark 3D	0.002	0.001	2.72	.007	Feature: landmark 3D	0.001	0.001	1.91	.057
Neural net: HMAX C2	–0.067	0.025	2.64	.008	Neural net: HMAX C2	–0.067	0.028	2.34	.019
Neural net: FaceNet	0.004	0.003	1.66	.097	Neural net: FaceNet	0.005	0.003	1.80	.072
Familiarity	0.077	0.022	3.54	<.001	Familiarity	0.079	0.024	3.26	.001

Note: The predictors of interest are in the first row (similarity rating: individual participants' pairwise dissimilarity rating between identities, trait distance: individual participants' euclidean distance of two identities based on trait ratings). B=unstandardized regression coefficient, SE=standard error, Z=Wald Z.

Study 4

Studies 2 and 3 found that when a participant regarded two individuals' personalities as more similar, estimated representations of their faces in the mind's eye of the participant became correspondingly more similar. In Study 4, we tested for a causal role of person knowledge in impacting facial representations by manipulating rather than measuring person knowledge. Participants first learned about novel individuals, who were either similar to one other (both trustworthy or both untrustworthy, Similar Pair) or dissimilar from one other in personality (one trustworthy and the other untrustworthy, Dissimilar Pair). They then proceeded to a reverse correlation task as the one in Study 3, which allowed us to assess representations of those individuals' faces. Conceptual similarity was measured for manipulation check via two indices – the explicit similarity rating and the euclidean trait distance.

Method

Participants. Three hundred and thirty individuals living in the US participated via MTurk. Eighteen participants were excluded for not following instructions, resulting in a final sample of 312 (66.70% male, 33.30% female; M_{age}=35.10 years, SD_{age}=9.85 years; 35.30% White, 41.70% Hispanic,

12.50% Black, 8.33% Asian, 2.24% other). To consider the difference in face-based trait impressions for each identity pair in our models as a potential confound (see *Analytic approach* below), a group of independent raters (n=47) were recruited to judge the 4 unfamiliar faces on 15 traits. Ten participants were excluded for not following instructions, resulting in 37 final participants (62.20% male, 37.80% female; M_{age}=37.90 years, SD_{age}=7.9 years 3; 24.30% Black, 46.00% Hispanic, 29.70% White).

Stimuli. To generate face stimuli, we first selected four identities that are not well known to people in the US. In order to approximately match the overall characteristics between faces used in Studies 1–3 and Study 4 (except for participants' familiarity with the faces), we selected four White males who are famous in another White-majority, industrialized country (i.e., Switzerland) who are in a similar age range and professions with to those in Studies 1–3: Didier Burkhalter (former politician), Pierre de Meuron (architect), Vincent Perez (actor), and Anatole Taubman (actor). As in Study 3, we first created each of the four individuals' faces by averaging three different images of the person using WebMorph (DeBruine, 2018). We then created a 50/50-blend morph between two identities for each pair of the total six pairs. Once we had the six morphs, we transformed them into vectors in the face space (Paysan et al., 2009) and created random variations of each of the morphed faces (i.e., 100 random face pairs varied on shape and 100 varied on color). As in Studies 2 and 3, for each identity pair we created 200 side-by-side face trials. See **Supplementary Figure 9** for details of stimulus preparation and reverse-correlation task procedures.

To vary the personality of identities, we used previously validated sentences that describe 2,375 social behaviors (Mende-Siedlecki & Havlicek, in preparation). In that work, participants had rated all 2,375 behaviors on trustworthiness, among other dimensions. We prepared four sets of sentences, selecting two sets from sentences rated as trustworthy (e.g., "Helped a neighbor fix his roof") and two sets from sentences rated as untrustworthy (e.g., "Took a few bills from the register at work"). Each set

consisted of 20 sentences. For each participant, two of these four sets would be randomly chosen and used to describe two novel identities in the experiment (i.e., trustworthy and trustworthy, untrustworthy and untrustworthy, trustworthy and untrustworthy).

Procedure. To directly manipulate participants' person knowledge of target identities, we used multiple stages: a prescreening stage, learning stage, conceptual-similarity task (manipulation check), and reverse correlation task. First, in the prescreening stage, participants were screened for their knowledge of the four target identities. Participants were asked to choose the face of each of the four identities in a series of multiple-choice questions, given each name. They were also asked to choose the occupation of each of the four identities. Only those who indicated that they did not know the answers to all questions were allowed to participate.

Each participant then learned about two of the four identities: either trustworthy and trustworthy (Similar Pair), untrustworthy and untrustworthy (Similar Pair), or trustworthy and untrustworthy (Dissimilar Pair). Participants were presented with a series of 20 slides (10 trials per identity), one at a time, with the face and a sentence describing the person's behavior below the face. The behavior was either trustworthy or untrustworthy, depending on the randomly assigned personality. For each target person, four face images were presented in randomized order across presentations. For each slide, which involved one face image and one behavioral sentence, participants were encouraged to "visualize the person doing the action described in the sentence as vividly as possible." The learning stage consisted of two blocks, each of which presented one of the two individuals' face image with their behaviors. Participants were then asked to choose the correct face corresponding to each of the two names they previously learned. They were also asked to choose the behavior that the two individuals were more likely to engage in; two options were given, one describing a trustworthy behavior and the other describing an untrustworthy behavior.

In the conceptual similarity task, participants answered an explicit pairwise conceptual similarity rating question (1 question per identity pair) and completed the personality trait rating task (15 traits per identity) about the two learned identities. The resulting conceptual similarity values served as a manipulation check that the Similar Pair condition successfully induced greater conceptual similarity than the Dissimilar Pair condition.

The reverse correlation task then followed the same procedures as Study 3, resulting in reverse-correlated images for each of the two identities for each participant. Across all participants, this resulted in a total of 312 images (2 identities × 6 identity pair conditions × 26 participants per condition). As in Study 3, we calculated the euclidean distance between the resulting two face vectors in face space for each participant, a value representing how similar the two faces were between two identities at the level of their objective features.

Analytic approach. We hypothesized that reverse-correlated images in the Similar Pair condition would exhibit a greater resemblance (i.e., higher levels of visual similarity) than those in the Dissimilar Pair condition. Thus, we conducted GEE multilevel regression analyses, where perceptual similarity was regressed onto Similar/Dissimilar Pair condition. As in the preceding reverse correlation studies (Studies 2–3), in each model we clustered data by identity pair, and included all visual similarity measures as additional predictors, so that the baseline visual similarity of the original facial photos was controlled for.

In Study 4, unlike in Studies 1–3, we randomly assigned person knowledge to identities to remove any effect of preexisting person knowledge related to target individuals. However, it remains possible that differences in the facial appearance of the four targets could, in theory, still exert some impact. We asked an independent group of raters (n=37) to make trait judgments of the four targets' faces. We assessed the same 15 traits used for measuring the personalities of the famous targets in the

preceding studies (adventurous, angry, anxious, assertive, cautious, cheerful, cooperative, depressed, dutiful, emotional, friendly, intellectual, self-disciplined, sympathetic, trustworthy) for use here with the four unfamiliar faces: "How [trait] is this person? Not at All [trait] 1–Extremely [trait] 7". Only those who were unfamiliar with all four targets could participate in the study (this was verified as part of the screening procedure). Trait ratings showed a high level of interrater agreement (ICC=0.70), thereby justifying our use of mean rating per target identity as a representative measure of its trait-related facial appearance. We calculated euclidean distances between trait vectors (15 coordinates) for every identity pair, thereby representing the extent to which any given pair of identities were similar/dissimilar in trait-related facial appearance, and included this distance as an additional covariate in our regression models. See **Supplementary Figure 10** for pairwise correlations between all covariates.

Results and Discussion

Confirming our manipulation, participants assigned to the Similar Pair condition deemed the pair of identities as having a higher conceptual similarity than those in the Dissimilar Pair condition, as indexed by both conceptual similarity ratings (n=156 each group; Welch's unequal variances t-test: $M_{Dissimilar}$=9.330, $SE_{Dissimilar}$=0.422, $M_{Similar}$=4.870, $SE_{Similar}$=0.146; t(310.00)=10.02, CI 95% [3.59,5.34], p<.001) and euclidean distance in trait space ($M_{Dissimilar}$=3.330, $SE_{Dissimilar}$=0.145, $M_{Similar}$=5.310, $SE_{Similar}$=0.095; t(268.02)=–11.48, CI 95% [-2.32,-1.64], p<.001).

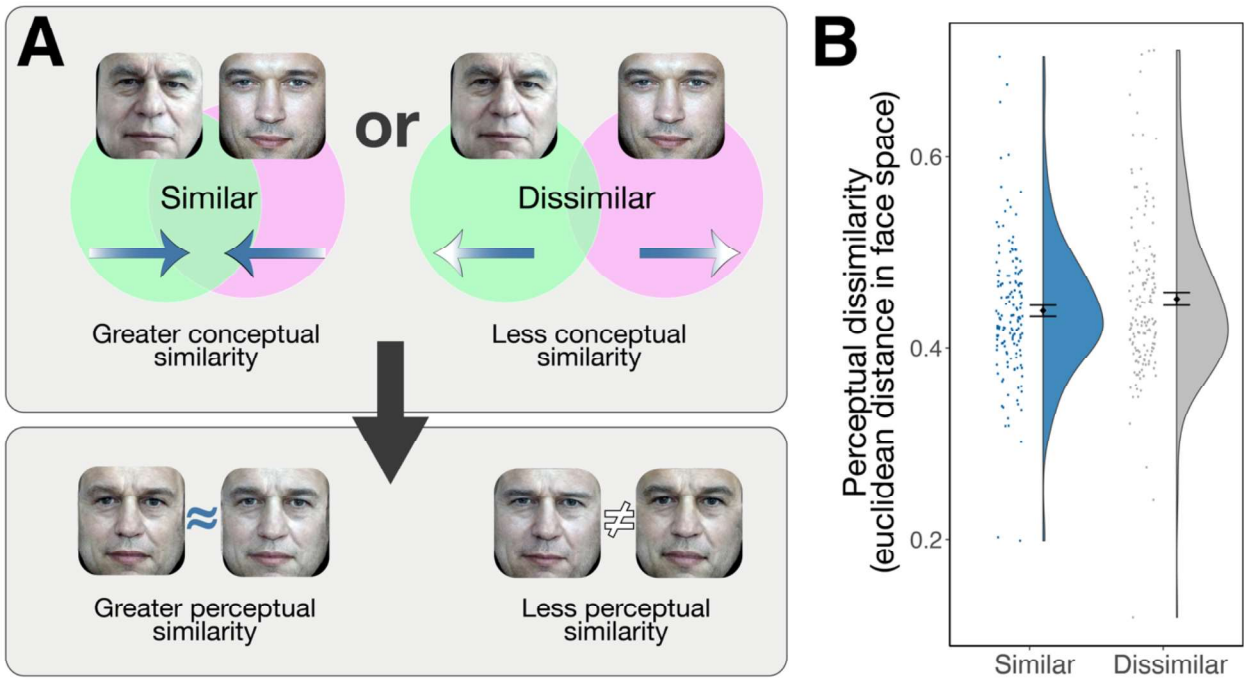

Figure 4. The analytic approach (A) and reverse correlation results (B) in Study 4. To test the causal effect of person knowledge on facial identity perception, we directly manipulated person knowledge and estimated participants' mental representations of the two individuals' faces (A). Participants learned about two novel identities whose personalities were either similar (i.e., both individuals were trustworthy, or both untrustworthy) or dissimilar to each other (i.e., one individual was trustworthy, and one was untrustworthy). They were assigned to one random face pair (two faces), both of which they found unfamiliar with. In the condition where two individuals' personalities were similar, participants' representations of their faces were more similar, compared to the condition where two individuals' personalities were dissimilar, as assessed by the euclidean distance between two reverse-correlated face vectors in a multidimensional space of faces (B). Actual analyses were conducted using GEE multilevel regressions. The violin plots represent the distribution of the perceptual dissimilarity between two target individuals' reverse-correlated images derived from each participant's responses. Each participant corresponds to a single datapoint. The black diamonds indicate the means of the perceptual dissimilarity metric in each condition. Error bars denote SEM.

A GEE regression predicting similarity in objective facial features from the person-knowledge condition (Similar Pair=–0.5, Dissimilar Pair=0.5) revealed the hypothesized effect, B=0.012, SE=0.005, 95% CI [0.002, 0.022], Z=2.41, p=.016 (**Figure 4B**, **Table 5**). Participants who learned that the identity pair had dissimilar personalities had more distinct perceptual representations of the two individuals' faces at the level of their objective features (M=0.451, SE=0.007), as compared to participants who learned that the pair had similar personalities (M=0.439, SE=0.006). Including as

covariates the visual similarity estimates of the original facial photos as well as similarity in trait-related appearance of the faces (along the 15 traits) led to identical results. Moreover, the full model that includes conceptual similarity explained the data better than a model that omits conceptual similarity ($\chi^2(1)= 5.79$, p=.016). Note that the covariate-only model (omitting conceptual similarity) also explained the data better than an intercept-only model ($\chi^2(4)=59.60$, p<.001), indicating that our visual model covariates faithfully captured visual information participants were utilizing for facial identity perception.

An alternative explanation for the effects may be that participants in the Similar Pair vs. Dissimilar Pair conditions differed in extraneous motivational or attentional processes. For example, when both identities had similar personalities, participants may not have been as motivated to discriminate between them; when the identities had dissimilar personalities, participants might have paid closer attention to their differences. If participants in the Similar Pair condition were less motivated or attentionally engaged, it is reasonable to expect that they would have exhibited worse performance in the learning stage, such as a diminished accuracy in matching identities to related behaviors, or would have been overall slower to respond on trials in either the learning stage or the reverse-correlation task. However, there was no evidence for this possibility. Participants in the Similar Pair condition did not differ from those in the Dissimilar Pair condition on any performance measures in either the learning stage or the reverse-correlation task, including when matching faces to behaviors in the learning stage (accuracy: $M_{Dissimilar}$=.73, $SE_{Dissimilar}$=.02, $M_{Similar}$=.77, $SE_{Similar}$=.02, t(305.75)=–1.53, CI 95% [–0.09,0.01], p=.127; response times: $M_{Dissimilar}$=5771.01 ms, $SE_{Dissimilar}$=559.30 ms, $M_{Similar}$ =5923.51 ms, $SE_{Similar}$=508.44 ms, t(305.24)=–0.20, CI 95% [–1635.05,1330.04], p=.840), when matching faces to names in the learning stage (accuracy: $M_{Dissimilar}$=.62, $SE_{Dissimilar}$=.03, $M_{Similar}$=.59, $SE_{Similar}$=.03, t(307.81)=0.69, CI 95% [–0.06,0.12], p=.491; response times: $M_{Dissimilar}$=5431.01 ms, $SE_{Dissimilar}$=996.90 ms, $M_{Similar}$=5739.62 ms, $SE_{Similar}$=517.97 ms, t(231.50)=–0.28, CI 95% [–2514.92,1897.69], p=.783), or

when selecting noise-imposed images in the reverse-correlation task (response times: $M_{Dissimilar}$=1327.92 ms, $SE_{Dissimilar}$=33.80 ms, $M_{Similar}$=1335.18 ms, $SE_{Similar}$=32.80 ms, t(309.72)=–0.15, CI 95% [–99.71,85.18], p=.877). These additional analyses cast doubt on the possibility that participants in two conditions differed in motivational or attentional processes, instead suggesting that genuine differences in the similarity/dissimilarity of person knowledge affected facial identity perception.

Thus, here we found that when two newly learned identities were similar in personality, the perceptual representations of their faces overlapped to a greater extent, compared to when the two newly learned identities were dissimilar in personality. These results suggest that person knowledge affected how similarly or dissimilarly participants perceived facial identities above and beyond any intrinsic similarities in the faces themselves. Crucially, because we randomly linked target identities' facial appearances with different personalities, these results suggest that the impact of person knowledge on facial identity perception is causal in nature: Our knowledge of others affects facial identity perception, independent of those individuals' facial features.

Table 5. Full Model Statistics in Study 4. Perceptual face-space distance was predicted from the Similar/Dissimilar Pair condition (i.e., the experimental condition that decided two target identities were either similar or dissimilar) in a GEE model.

Predictor	B	SE	Z	p
Conceptual: Similar/Dissimilar Pair	0.012	0.005	2.41	.016
Feature: head rotation	–0.093	0.065	1.44	.151
Feature: landmark 2D	<0.001	<0.001	1.24	.214
Feature: action unit	0.005	0.003	1.99	.046
Face trait distance	–0.012	0.003	4.45	<.001

Note: The predictor of interest is in the first row (a binary variable indicating whether two individuals' personalities were manipulated to be perceived similar vs. dissimilar).

General Discussion

Across four studies, we found that knowledge of others' personalities shapes the perception of their facial identities. We assessed the effect of person knowledge by mapping measures across conceptual and perceptual levels. Using mousetracking, we found that perceivers' own unique conceptual similarity among identities predicted biases in perceiving facial identities, even while acknowledging intrinsic physical similarity among faces (Study 1). Using converging reverse correlation techniques, we found that representations of individual faces became perceptually more similar when any two individuals were deemed to be more similar in their personalities (Studies 2 and 3). Finally, we showed that participants who thought that two identities had a similar vs. dissimilar personality had representations of faces that were correspondingly similar vs. dissimilar, implicating person knowledge's causal role on how facial identity is represented (Study 4). Together, converging evidence across the process of facial identity perception (mousetracking) and estimated perceptual representations (reverse correlation) suggests that person knowledge has the power to dynamically shape facial identity perception, biasing it toward alternate identities despite the fact that those identities lack any physical resemblance.

Previous research has shown that conceptually related face primes or contextual cues can facilitate successful face recognition (Bruce, 1983; Bruce & Valentine, 1986), and such malleability from conceptual processes is consistent with longstanding models of face recognition (Bruce & Young, 1986; Burton et al., 1999; Burton et al., 1990). However, prior work has left unaddressed the unique perceptual biases that may arise when overlapping person knowledge causes the faces of ostensibly unrelated individuals to be perceived more similarly. This premise is consistent with connectionist models of face recognition and is theoretically predicted by newer person perception models (Freeman & Ambady, 2011; Freeman et al., 2020). It is also consistent with recent findings highlighting the role of conceptual learning in face recognition (Gordon & Tanaka, 2011; Yovel et al., 2012). The current

findings bolster these face perception models, providing novel evidence that the process of perceiving facial identity is dynamically constructed not only by the visual processing of a face but also our own social-conceptual associations. The results add to a growing body of research demonstrating a variety of social-conceptual impacts on face perception. Studies have shown that one's stereotypes affect social category perception from faces (Johnson, Freeman, & Pauker, 2012; Stolier & Freeman, 2016), one's emotion concepts affect the perception of facial expressions (Brooks et al., 2019; Brooks & Freeman, 2018; Carroll & Young, 2005; Lindquist, Gendron, Barrett, & Dickerson, 2014), and one's personality concepts affect trait impressions from faces (Oh, Martin, & Freeman, under review; Stolier et al., 2020; Stolier, Hehman, Keller, Walker, & Freeman, 2018). Moreover, neural decoding studies suggest that these social-conceptual impacts reach relatively early levels in the perceptual representation of faces (Brooks et al., 2019; Stolier & Freeman, 2016).

The present studies are not without their limitations. While the mousetracking and reverse correlation approaches aim to provide a window into the perceptual process of recognizing facial identity and how facial identity is perceptually represented, neither of these measures can be considered a "pure" measure of perception. The potential roles of attentional or post-perceptual decision processes also cannot be excluded. Particularly for the mousetracking data, it is possible that conceptual influences occur at higher levels of processing than are hypothesized here (De Falco, Ison, Fried, & Quiroga, 2016). Combining our current approach with neural decoding techniques could help better identify at which levels of representation these effects manifest (Freeman et al., 2018). We should also note that, although our focus here is on person knowledge in the form of impressions of others' personality traits, we would argue that any form of person knowledge (e.g., occupation, situational information about a person) would be a candidate for the effects observed. Although it is plausible that the pairwise conceptual ratings were additionally influenced by these other factors (as they indexed person

knowledge globally), our direct measurement of participants' impressions of the 15 personality traits and our manipulation of these impressions are unlikely to have been influenced by these factors. Thus, while a trial involving Bill Clinton and Vladimir Putin, for example, in theory may have led to stronger perceptual similarity than a trial involving Justin Bieber and Vladimir Putin due to extraneous person-knowledge factors such as a congruent occupation (e.g., politics), our direct measurement of the 15 personality trait impressions and manipulation of those impressions provide strong evidence for the specific impact of personality impressions on perception. Nevertheless, as just mentioned, these additional forms of person knowledge are certainly consistent with our overall theoretical account, and future work could examine these additional factors' impact on perception directly.

A caveat is also due regarding any pure delineation between the conceptual knowledge of targets' personality and the perceptual processing of their facial appearance. These are not fully independent, as facial appearance can have complex influences on one's personality just as one's personality can have complex influences on facial appearance (Zebrowitz & Collins, 1997; Zebrowitz, Collins, & Dutta, 1998). With respect to measurement, it is plausible that participants' conceptual similarity ratings (person similarity and trait ratings) may have been implicitly influenced by facial appearance, raising the possibility of a bottom-up perceptual confound. However, the nature of our statistical analyses ensured that there were meaningful individual differences in conceptual similarity that manifested as corresponding differences in identity perception; such findings cannot be explained by bottom-up perceptual confounds in facial appearance that would be identical across participants. Moreover, the inclusion of multiple visual similarity estimates as covariates in the regression analyses additionally casts doubt on the idea that facial appearance may have confounded the effects of interest. Finally, in the only study involving unfamiliar faces (where trait-related facial appearance estimates are valid, unlike with familiar faces), conceptual knowledge was found to affect perceptual representations

of faces even when statistically controlling for trait-related facial appearance. Together, our results suggest that the effects of person knowledge on facial identity perception cannot be explained by bottom-up perceptual confounds in facial appearance alone.

We also expect the observed effects to be bound by context (e.g., biasing effects toward identities that are currently accessible due to context) and, in the case of the mousetracking study, to be temporary. Past mousetracking studies have suggested that even temporary biases during the early moments of face perception may have lingering downstream social consequences (for review, Freeman & Johnson, 2016). Indeed, previous research suggests that it is possible that the biasing effects of person knowledge on facial identity perception may bear downstream consequences. Impressions of a person's personality are transmitted to a novel identity when the two individuals' faces resemble one another, as revealed via behavioral (e.g., impressions of trustworthiness, decisions to trust) and neural measures (e.g., amygdala activation) (FeldmanHall et al., 2018; Verosky & Todorov, 2010, 2013). When a perceiver's person knowledge biases a face to appear more similar to the face of an unrelated individual, such biased resemblance in appearance could in turn affect how we feel and think about the new individual.

The current findings highlight both the benefit and costs of the impact of person knowledge on social behavior. Social judgments are an effortless, largely uncontrollable, and highly efficient process. Perceiving faces of different individuals similarly when they share similar personality may facilitate an efficient storage of social information and promote adaptive interpersonal behavior (e.g., a quick decision to help someone or not when their face resembles a close friend). On the other hand, the perceptual overlap between identities induced by conceptual similarity between them could cause unwanted confusion due to the diminished precision in facial representation. Thus, although we argue that the effects of person knowledge on facial identity perception are a byproduct of more domain-

general interactive visual processing (Freeman et al., 2020) (and such processing affords clear evolutionary advantages, e.g., Gilbert & Li, 2013), the effects demonstrated here are likely to function adaptively in some contexts while maladaptively in others. Future research could investigate to what extent the social-conceptual scaffolding of facial identity perception may translate into cognitive, affective, and behavioral consequences out in the social world, for better or for worse.

Data and Materials Availability

All data, stimuli, and analysis code are made available via Open Science Framework: https://osf.io/rcmyh/.

Acknowledgements

The authors thank Ian Ferguson, Loris Jeitziner, Yanzi Huang, Roshni Lulla, and Inshil Paik for their assistance. This work was supported in part by research grant BCS-1654731 (J.B.F.).

Reference

Ahumada Jr, A., & Lovell, J. (1971). Stimulus features in signal detection. *The Journal of the Acoustical Society of America, 49*(6B), 1751-1756.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). *Openface 2.0: Facial behavior analysis toolkit.* Paper presented at the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology, 28*(1), 333-361. doi:10.1080/10463283.2017.1381469

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences, 116*(32), 15861-15870. doi:10.1073/pnas.1816408116

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour, 2*(8), 581-591.

Bruce, V. (1979). Searching for politicians: An information-processing approach to face recognition. *Quarterly Journal of Experimental Psychology, 31*, 373-395.

Bruce, V. (1983). Recognizing faces. *Philosophical Transactions of the Royal Society of London, Series B, 302*, 423–436.

Bruce, V., & Valentine, T. (1986). Semantic priming of familiar faces. *The Quarterly Journal of Experimental Psychology Section A, 38*(1), 125-150. doi:10.1080/14640748608401588

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science, 23*(1), 1-31. doi:10.1207/s15516709cog2301_1

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*(3), 361-380. doi:10.1111/j.2044-8295.1990.tb02367.x

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *Quarterly Journal of Experimental Psychology, 58*(7), 1173-1197. doi:10.1080/02724980443000539

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO Personality Inventory Revised–(NEO-PIR) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

De Falco, E., Ison, M. J., Fried, I., & Quiroga, R. Q. (2016). Long-term coding of personal and universal associations underlying the memory web in the human brain. *Nature Communications, 7*(1), 1-11.

DeBruine, L. (2018). debruine/webmorph: Beta release 2 (Version v0.0.0.9001). Retrieved from https://webmorph.org/

Dotsch, R. (2015). rcicr: Reverse correlation image classification toolbox (Version 0.3.2) [R package].

Dotsch, R., & Todorov, A. (2011). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562-571. doi:10.1177/1948550611430272

Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision, 2*(1), i-i.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A. T., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences, 115*(7), E1690–E1697.

Folstein, J. R., Palmeri, T. J., Van Gulick, A. E., & Gauthier, I. (2015). Category learning stretches neural representations in visual cortex. *Current Directions in Psychological Science, 24*(1), 17-23. doi:10.1177/0963721414550707

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science, 27*(5), 315–323.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247-279. doi:10.1037/a0022327

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20*(5), 362-374. doi:10.1016/j.tics.2016.03.003

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

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. S. (2018). The neural representational geometry of social perception. *Current opinion in psychology, 24*(1), 83-91. doi:10.1016/j.copsyc.2018.10.003

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nat Rev Neurosci, 14*(5), 350-363. doi:10.1038/nrn3476

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe, 7*(1), 7-28.

Gordon, I., & Tanaka, J. W. (2011). The role of name labels in the formation of face representations in event-related potentials. *British Journal of Psychology, 102*(4), 884-898. doi:10.1111/j.2044-8295.2011.02064.x

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review, 14*(6), 1085-1089. doi:Doi 10.3758/Bf03193095

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233. doi:10.1016/S1364-6613(00)01482-0

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

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577-596.

Kidder, C. K., White, K. R., Hinojos, M. R., Sandoval, M., & Crites Jr, S. L. (2018). Sequential stereotype priming: A meta-analysis. *Personality and Social Psychology Review, 22*(3), 199-227.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci, 2*(4), 1–28. doi:10.3389/neuro.06.004.2008

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*(2), 375-387. doi:10.1037/a0035293

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology, 37*(5), 793-805.

Mandler, G. (1980). Recognizing - the Judgment of Previous Occurrence. *Psychological Review, 87*(3), 252-271. doi:Doi 10.1037/0033-295x.87.3.252

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science, 28*(2), 209-226. doi:10.1207/s15516709cog2802_4

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.

Mende-Siedlecki, P., & Havlicek, L. (in preparation). *The Delaware behavior database: A set of social behaviors and corresponding norming data*.

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in Cognitive Sciences, 22*(9), 794-809. doi:10.1016/j.tics.2018.06.006

Oh, D., Martin, J. D., & Freeman, J. B. (under review). Personality across world regions predicts variability in the structure of face impressions.

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

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). *Deep face recognition.* Paper presented at the Proceedings of the British Machine Vision Conference (BMVC).

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). *A 3D face model for pose and illumination invariant face recognition.* Paper presented at the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in Cognitive Sciences, 22*(6), 471–472.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). *FaceNet: A unified embedding for face recognition and clustering.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 411-426. doi:10.1109/TPAMI.2007.56

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience, 19*(6), 795-797. doi:10.1038/nn.4296

Stolier, R. M., & Freeman, J. B. (2017). A neural mechanism of social categorization. *Journal of Neuroscience, 37*(23), 5711-5721.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour, 4*, 361–371. doi:10.1038/s41562-019-0800-6

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. doi:10.1073/pnas.1807222115

Sunday, M. A., & Gauthier, I. (2018). Face expertise for unfamiliar faces: A commentary on Young and Burton's "Are we face experts?". *Journal of Expertise, 1*(1), 35–41.

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. doi:10.1016/j.cognition.2012.12.001

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*(6), 401–409.

Tiddeman, B., Burt, D., & Perrett, D. (2001). Computer graphics in facial perception research. *IEEE Computer Graphics and Applications, 21*(5), 42-50.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161-204. doi:10.1080/14640749108400966

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science, 21*(6), 779-785. doi:10.1177/0956797610371965

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology, 49*(4), 661-669. doi:10.1016/j.jesp.2013.02.004

Walker, M., & Keller, M. (2019). Beyond attractiveness: A multimethod approach to study enhancement in self-recognition on the Big Two personality dimensions. *Journal of Personality and Social Psychology, 117*, 483-499. doi:10.1037/pspa0000157

Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision, 9*(11), 1-13. doi:10.1167/9.11.12

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*(3), 395–412.

Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. *Annual Review of Psychology, 43*(1), 473-504.

Winograd, E., & Riversbulkeley, N. T. (1977). Effects of Changing Context on Remembering Faces. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 397-405. doi:Doi 10.1037/0278-7393.3.4.397

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences, 22*(2), 100-110. doi:10.1016/j.tics.2017.11.007

Young, A. W., Flude, B. M., Hellawell, D. J., & Ellis, A. W. (1994). The nature of semantic priming effects in the recognition of familiar people. *British Journal of Psychology, 85*(3), 393-411. doi:10.1111/j.2044-8295.1994.tb02531.x

Young, A. W., Hellawell, D., & De Haan, E. H. F. (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. *The Quarterly Journal of Experimental Psychology Section A, 40*(3), 561-580. doi:10.1080/02724988843000087

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 285-289. doi:10.1037/a0027077

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review, 1*(3), 204-223.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and Social Psychology Bulletin, 24*(7), 736-749.