Social Cognition, Vol. 40, No. 6, 2022, pp. 549–579

EXPLORING THE REPRESENTATIONAL STRUCTURE OF TRAIT KNOWLEDGE USING PERCEIVED SIMILARITY JUDGMENTS

Minjae Kim, Liane Young, and Stefano Anzellotti
Boston College

A large body of past work has sought to identify the underlying dimensions that capture our trait knowledge of other people. However, the *importance* of particular traits in determining our overall impressions of others is not well understood, and different traits may be fundamental for impressions of famous versus unfamiliar people. For instance, we may focus on competence when evaluating a famous person, but on trustworthiness when evaluating a stranger. To examine the structure of overall impressions of famous people and of unfamiliar people, we probed the contributions of 13 different trait judgments to perceived similarity judgments. We found that different sets of traits best predicted perceived similarity between famous people versus between unfamiliar people; however, the relationship between each trait and perceived similarity generalized to some extent from famous people to unfamiliar people, suggesting a degree of overlap in the structure of overall impressions.

Keywords: person perception, social cognition, trait inference, similarity judgments

---

The data and analysis code for this research and supplementary materials are available at https://osf.io/kqc2h/?view_only=1e2116e04ea04b19accede3330e412ba).

The authors would like to thank Tony Chen, Isaac Handley-Miner, Ryan McManus, Stella Si, the Boston College Morality Lab, and the Social and Cognitive Computational Neuroscience lab for their helpful feedback. This work was supported by the John Templeton Foundation (61061 to Liane Young), the National Science Foundation (1627157 to Liane Young and CAREER Grant 1943862 to Stefano Anzellotti), and the Boston College Research Incentive Grant (to Liane Young and Minjae Kim).

Minjae Kim, Liane Young, and Stefano Anzellotti conceived of and designed the study; Minjae Kim collected and analyzed the data; Minjae Kim, Liane Young, and Stefano Anzellotti interpreted the data and wrote the manuscript.

The authors declare no competing interests.

Address correspondence to: Minjae Kim, Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA 02467. E-mail: minjae.kim@bc.edu

INTRODUCTION

How is our knowledge of other people organized? According to dimensional theories of social cognition, our knowledge of others' psychological characteristics, such as mental states and traits, can be represented by coordinates within a space defined by multiple evaluative dimensions (Bach & Schenke, 2017; Cuddy et al., 2008; Tamir & Thornton, 2018; Thornton et al., 2019). For instance, while faces can elicit many different trait inferences, variance in face-based trait inference is well-described by two underlying dimensions, called valence (approximated by judgments of trustworthiness) and dominance (approximated by judgments of social dominance; Oosterhof & Todorov, 2008). Traits form a particularly important part of person knowledge: They are inferred characteristics that vary across individuals but remain relatively stable within an individual over time (Allport & Odbert, 1936). As such, trait knowledge enables perceivers to tailor an understanding of behaviors to specific individuals, and to generate predictions about possible future actions and reactions across contexts (Gerstenberg et al., 2018; Kryven et al., 2016; Wu et al., 2018; see Bach & Schenke, 2017, and Tamir & Thornton, 2018, for reviews on the use of social knowledge for prediction). For instance, the position of a face within the valence–dominance space described above can be used to accurately predict threat evaluations, which have adaptive significance (Oosterhof & Todorov, 2008). Understanding the representational structure of trait knowledge, then, is key to understanding how people interpret and predict behavior.

A large body of psychological research has sought to identify the underlying dimensions that capture perceivers' trait knowledge of others. Thornton and Mitchell (2018) describe four such dimensional theories of person perception that have been influential in the literature: (1) the five-factor model of personality, which consists of openness, conscientiousness, extraversion, agreeableness, and neuroticism (Goldberg, 1990; McCrae & Costa, 1987); (2) the stereotype content model, which consists of warmth and competence (Fiske et al., 2002); (3) the two-factor model of mind perception, which consists of agency and experience (Gray et al., 2007); and (4) the two-factor model of face perception, which consists of trustworthiness and dominance (Oosterhof & Todorov, 2008). Each of these theories was originally developed to account for specific phenomena: judgments of trait terms, intergroup affect, mind attribution, and face evaluation, respectively.

These theories have been tested in a common framework by harnessing the multidimensionality of fMRI data. Thornton and Mitchell (2018) scanned participants while they made social judgments (e.g., "loves to solve difficult problems"; "enjoys spending time in nature") about famous people that had been selected to span a variety of traits. Neural pattern responses to famous people in this task were predicted by each of the four aforementioned theories of person perception (Fiske et al., 2002; Goldberg, 1990; Gray et al., 2007; McCrae & Costa, 1987; Oosterhof & Todorov, 2008), and by a three-factor synthetic model, produced by applying principal component analysis to the four extant theories. In addition, the three-factor synthetic model outperformed all four extant theories in neural pattern reconstruction. These findings show that (1) dimensional theories of social cognition may partially describe

the informational basis of mentalizing; (2) these theories can generalize beyond their original contexts (of personality, intergroup affect, mind attribution, and face evaluation); and (3) pooling dimensions across inference contexts allows researchers to capture a greater proportion of the reliable variance in neural responses to famous people. In all, extant dimensional theories of person perception seem to be viable accounts of how perceivers represent other people during mentalizing.

Despite extensive previous research on the structure of trait knowledge, the importance of each individual trait in determining *overall impressions* of others is not as well understood. In addition, the traits that play a predominant role in determining overall impressions of famous or familiar people may be different from the traits that are fundamental for overall impressions of unfamiliar people. Previous fMRI studies have revealed that distinct brain regions are engaged in the representation of famous, familiar, and unfamiliar individuals (Gorno-Tempini & Price, 2001; Grabowski et al., 2001; Ramon & Gobbini, 2018), suggesting the possibility that representations of famous people are organized differently than representations of unfamiliar people.

It is difficult to investigate the importance of different traits for overall representations of people using fMRI responses alone. Readout mechanisms are needed to convert neural representations of traits into behavioral judgments (Pagan et al., 2016; Park et al., 2014). As such, even if a dimension explains a large amount of variance in neural responses to people, it may still contribute to a lesser degree to behavioral judgments of people. Behavioral studies can therefore make unique contributions to the investigation of the structure of person representations.

Previous behavioral studies have largely relied on principal component analysis (PCA) to identify the key dimensions that capture variance in trait judgments. PCA is a simple and elegant technique that identifies dimensions that account for most variance in a dataset, and as such effectively uncovers a "compressed" description of the dataset. It has been used successfully to identify lower-dimensional representational spaces that capture variance in perceivers' judgments of people along a set of specified traits (McCrae & Costa, 1987; Thornton & Mitchell, 2018). However, there is no guarantee that the dimensions that explain the most variance across trait judgments (and by extension, the *traits* that best approximate the content of these dimensions) *also* contribute the most to overall representations of people. When a participant is asked to evaluate a person along a specific trait (e.g., "How open to experience is this person?"), such judgments do not carry information about the importance of that particular trait for the overall representation of that person. To examine the importance of different traits for overall representations, we surveyed how trait judgments contribute to perceived similarity judgments between target people.

USING PERCEIVED SIMILARITY TO CHARACTERIZE TRAIT KNOWLEDGE

In this study, we investigated the importance of 13 different traits in determining: (1) overall representations of famous people, and (2) overall representations of

unfamiliar people who were described as performing a single behavior. Specifically, we aimed to identify the traits that contribute most to perceived similarity ratings between pairs of target people (collected by asking: "How similar are these two people?").

The perceived similarity approach has previously been used to test the "summed state" hypothesis of person representations (Thornton et al., 2019): Thornton and colleagues showed that both perceived similarity ratings and neural pattern similarities were better predicted by a model that reflects how frequently targets experience mental *states*, rather than by an optimized model of *traits*. The trait model, however, was still a robust predictor of similarity, and explained unique variance beyond the summed state model, indicating that traits still play a significant role in person representation.

In the current work, we examined the contributions of 13 traits (collated from extant theories of person perception by Thornton & Mitchell, 2018) to overall representations of people. To do this, we tested whether pairwise differences between targets along individual traits (i.e., trait distances) can predict pairwise holistic similarity ratings. For example, if inferences of *openness to experience* are important in determining overall representations of people, then the distance between two targets in terms of openness ratings should predict how (dis)similar the two targets are rated to be overall. Importantly, surveying how trait distances predict holistic similarity is a way to implicitly assess how perceivers prioritize and integrate multiple trait judgments to form overall representations. Additionally, the traits that perform best in predicting holistic similarity may not necessarily be ones that have traditionally been considered together; that is, the top-performing traits may cut across different theories that have been proposed for specific contexts of social inference.

TRAIT KNOWLEDGE ACROSS INFERENCE CONTEXT

We have discussed previous work that investigated representations of famous people (Thornton & Mitchell, 2018). Other studies have tested how we update representations of unfamiliar people, given information about their behaviors (e.g., Kim et al., 2021; Mende-Siedlecki et al., 2013). These paradigms involve different kinds of inference, and may elicit different person representations. When participants make social judgments about a famous person, they might draw on behavioral observations across different contexts. They might also have additional knowledge about them acquired through language (e.g., by reading a newspaper article). By contrast, participants exposed to an unfamiliar person described as performing a single behavior have access to impoverished information for trait inferences, and they may represent that person differently.

In addition, the dimensionality of person representations could itself change as a function of the amount and type of evidence available. A higher-dimensional representation would require estimating a larger number of coordinates, and thus would require a correspondingly larger amount of data in order to obtain robust estimates. Considering this, the dimensionality of perceivers' representations of

other people might be adaptive, adjusting optimally to the amount of information we have about a particular individual (e.g., representations of strangers may be lower-dimensional than representations of known individuals).

In order to study person representations across different inference contexts, we conducted perceived similarity analyses on two datasets: ratings of famous people (collected by Thornton & Mitchell, 2018), and ratings of unfamiliar people who performed a single behavior. For each domain (famous people and unfamiliar people), we tested how well pairwise trait distances predicted pairwise holistic similarity. We also tested whether the mappings between trait distance and holistic similarity generalized across the two domains. We found that distinct subsets of traits best predicted holistic similarity between famous people versus between unfamiliar people. However, the relationship between each trait and holistic similarity generalized to some extent from famous people to unfamiliar people, suggesting a degree of overlap in representational structures across inference contexts. As compared to trait ratings of famous people, trait ratings of unfamiliar people were more intercorrelated, and they were largely driven by valence (positivity or negativity). However, removing the influence of valence information revealed that a reliable higher-dimensional structure was present even in first impressions.

METHODS

SET OF EXAMINED TRAITS

Thirteen traits tested in a previous study of neural pattern activity during mentalizing (Thornton & Mitchell, 2018) were examined in the current study. Thornton and Mitchell (2018) took 11 of these from four extant theories of person knowledge and face perception: warmth and competence from the stereotype content model (Fiske et al., 2002); agency and experience from the two-factor model of mind perception (Gray et al., 2007); trustworthiness and dominance from the two-factor model of face perception (Oosterhof & Todorov, 2008); and the Big 5 personality dimensions, openness, conscientiousness, extraversion, agreeableness, and neuroticism (Goldberg, 1990; McCrae & Costa, 1987). Intelligence and attractiveness were also included for being widely discussed features in person knowledge (Thornton & Mitchell, 2018).

TRAIT RATINGS OF UNFAMILIAR PEOPLE: OVERVIEW

There were two rounds of data collection for trait ratings of unfamiliar people. In the first round of data collection, participants rated a set of *nameless and faceless* target people, who were each described as performing a single behavior. While participants were instructed to give trait ratings of unfamiliar *people* based on their behaviors, participants may have instead rated the *behaviors* themselves, as the targets were not highly personified. Thus, we conducted a conceptual replication study where *named and pictured* target people were described as performing a single behavior.

Following these two rounds of data collection, we assessed whether trait ratings of unfamiliar people with and without names and faces were comparable. We found that the two datasets were highly concordant (see Results). Therefore, for downstream data collection (of holistic similarity ratings) and analyses, we focused on unfamiliar targets *without* names and faces.

BEHAVIOR STIMULI ASSOCIATED WITH UNFAMILIAR PEOPLE

Three hundred single-sentence descriptions of behaviors were taken from a previous study of neural activity during impression updating (Kim et al., 2021; stimuli adapted from Mende-Siedlecki et al., 2013). Of these, 120 behaviors were positive/moral (e.g., "spent a Saturday volunteering at a soup kitchen"), 120 were negative/immoral (e.g., "lost their temper at the barista"), and 60 were neutral/morally irrelevant (e.g., "walked down a sidewalk in town"). All behavior stimuli were pretested to verify valence (positivity or negativity) and moral relevance (Kim et al., 2021).

TRAIT RATINGS OF UNFAMILIAR PEOPLE (WITHOUT NAMES AND FACES)

Participants were recruited through Amazon Mechanical Turk (MTurk) to rate a set of 60 unfamiliar people on a single trait. Five surveys were administered for each trait, to present all 300 behavior stimuli. We aimed to recruit approximately 30 participants for each of 65 surveys (13 traits × 5 surveys per trait); of the 2,059 participants that were recruited in total, 74 were excluded for failing attention checks or for being non-native speakers of English, resulting in a final sample of 1,985 participants (995 female, 958 male, 6 nonbinary/other participants; age $M = 37.2$, $SD = 11.2$).

For each item, participants were asked to imagine someone who performed one behavior (e.g., "Imagine a person who spent a Saturday volunteering at a soup kitchen"). Participants then rated that person along the specified trait, on a scale from 1 to 7 (e.g., "Please rate the openness to experience of this person"). A short description of the trait was provided at the beginning of each survey (see supplementary materials, p. 16, for full participant instructions).

TRAIT RATINGS OF UNFAMILIAR PEOPLE (WITH NAMES AND FACES)

A new set of participants was recruited through MTurk to rate a set of 60 unfamiliar people (30 female, 30 male) on a single trait. Five surveys were administered for each trait, to present all 300 behavior stimuli. We aimed to recruit approximately 10 participants for each of 65 surveys (13 traits × 5 surveys per trait); of 700 total participants, 46 were excluded for failing attention checks or for being non-native speakers of English, resulting in a final sample of 654 participants (298 female, 351 male, 3 nonbinary/other participants; age $M = 39.4$, $SE = 12.0$).

Each target person was given a name and was represented by a picture of an emotionally neutral face from the Karolinska Directed Emotional Faces set (Lundqvist et al., 1998). Each person was described as performing one behavior (e.g., "Andrew spent a Saturday volunteering at a soup kitchen"). Participants were asked to rate each person on the specified trait, on a scale from 1 to 7 (e.g., "Please rate the openness to experience of this person"). A short description of the trait was provided at the beginning of each survey. Across participants, target identity was counterbalanced with behavior valence (e.g., half of participants learned about Andrew performing a positive behavior, and half of participants learned about him performing a negative behavior).

HOLISTIC SIMILARITY RATINGS FOR PAIRS OF UNFAMILIAR PEOPLE

Holistic similarity ratings were collected for 900 randomly chosen pairs of unfamiliar people (out of $C[300, 2] = 44{,}850$ possible pairs). As discussed above, we only collected holistic similarity ratings for unfamiliar targets *without* names and faces, because (1) the inclusion of names and faces did not impact trait ratings (see Results), and (2) participants may overweigh facial similarity in their holistic similarity ratings if pictures of faces are presented.

A new set of participants was recruited through MTurk to rate 60 stimulus pairs. Fifteen surveys were administered to present all 900 stimulus pairs. We aimed to recruit approximately 5 participants for each survey; of 79 total participants, 4 were excluded for failing attention checks or for being non-native speakers of English, resulting in a final sample of 75 participants (38 female, 36 male, 1 nonbinary / other participants; age $M = 30.1$, $SD = 12.8$).

For each stimulus pair presented in the survey, participants were asked to imagine one person performing the first behavior, and another person performing the second behavior; then, participants rated how similar the two people were, on a scale from 0 (extremely dissimilar) to 100 (extremely similar). For example: "Imagine that one person spent a Saturday volunteering at a soup kitchen. Imagine that another person lost their temper at the barista. How similar are these two people?"

Following data collection, pairwise holistic similarity ratings were reflected, such that higher ratings indicated greater dissimilarity (distance) between the two targets.

All participants in the above studies provided informed consent and were compensated for their time; for full participant demographics please see supplementary materials (p. 18).

RATINGS OF FAMOUS PEOPLE

Trait ratings and pairwise holistic similarity ratings of 60 famous people (e.g., Amelia Earhart, Bruce Lee, George W. Bush) were taken from Thornton and Mitchell (2018). Thornton and Mitchell (2018) collected ratings of the 60 targets on each of the 13 traits from an online sample ($N = 869$). Each participant rated the entire

set of 60 targets on a single trait. A short description of the relevant trait was provided. Participants gave their ratings on a continuous line scale from 1 to 7 with anchors appropriate to the trait. In addition, a separate set of participants gave holistic similarity ratings for every pair of targets (Thornton & Mitchell, 2018). Of the $C[60, 2] = 1{,}770$ pairwise holistic similarity ratings, we randomly selected and retained 900 for further analysis in the current study, to match the number of holistic similarity ratings that were collected for unfamiliar people.

TRAIT DISTANCE CALCULATION

For each stimulus pair for which we had holistic similarity ratings (900 pairs of unfamiliar people, 900 pairs of famous people), we computed 13 pairwise trait distances. Trait distance was calculated as the absolute difference between the average trait rating for one target and the average trait rating for the other target.

PREDICTING HOLISTIC SIMILARITY USING TRAIT DISTANCE

All analyses were conducted in R (R Core Team, 2013). For each domain (unfamiliar people and famous people), we fit 13 single-variable linear models (ordinary least-squares) to predict pairwise holistic similarity (reflected) using pairwise trait distance. For example, one model predicted holistic similarity between pairs of unfamiliar people as a function of their distance along openness. We used the Holm-Bonferroni method to correct p values for the models. For each domain, we also fit a cumulative linear model, where all 13 trait distances were used to predict pairwise holistic similarity, to explore how much of the variance in holistic similarity could be explained by extant theories of person perception. For cumulative models, partial correlations were calculated between each trait distance and holistic similarity.

For the domain of famous people, we also tested whether associations between holistic similarity and trait distance would be robust to adding biographical information as covariates. For each pair of famous people, we coded whether or not the targets shared the same gender, race, nationality, and industry (arts, athletics, business, media, politics, sciences), based on Wikipedia entries (entering NAs where information was not available). These four covariates were added to all models predicting holistic similarity based on trait distance.

COMPARING WITHIN-DOMAIN AND CROSS-DOMAIN PREDICTIVE PERFORMANCE

For each linear model in each domain, five-fold cross-validation was used to examine within-domain predictive performance and cross-domain predictive performance. For instance, models trained on the unfamiliar people data were used to predict: (1) holistic similarity for held-out pairs of unfamiliar people (*within-domain generalization*), and (2) holistic similarity for pairs of famous people (*cross-domain generalization*).

To do this, we randomly split the rating data in each domain into five folds, iterating through each fold as the test (held-out) set. Standardization of all variables was conducted separately for training and test sets. Linear models that were fitted to the training set in one domain were used to predict: (1) holistic similarity values in the same domain's test set, and (2) holistic similarity values in the other domain's test set. For example, one model, which regressed holistic similarity onto distance along openness, was trained on folds 1–4 of the unfamiliar people data; this model was then used to predict holistic similarity as a function of distance along openness in fold 5 of the unfamiliar people data, and in fold 5 of the famous people data.

The following measures of predictive performance were averaged across the five folds: coefficient of determination (CoD; calculated as 1 – sum of squares error/sum of squares total), root mean squared error (RMSE), and mean absolute error (MAE).

This five-fold cross-validation procedure was repeated with the cumulative models in each domain, where all 13 trait distances were used to predict holistic similarity.

These performance measures allowed us to examine: (1) the importance of different traits for explaining holistic similarity, and (2) whether there were correspondences in how traits related to holistic similarity across inference contexts.

CORRELATION STRUCTURES

We next examined whether the two domains—unfamiliar people and famous people—differed in terms of collinearity between trait ratings.

For the set of unfamiliar people, and for the set of famous people, we generated a correlation matrix that plotted the Pearson's correlation coefficient for all pairwise combinations of the 13 trait ratings. Then, we conducted chi-squared tests of whether the Fisher-transformed correlation matrices were significantly different, using the cortest.mat function in R.

RELIABILITY OF CORRELATION STRUCTURES

Next, we examined the reliability of the correlation structures for the two domains, as any differences in intercorrelatedness may be due to greater noise in one dataset. For each dataset, we randomly generated a subset of 60 stimuli (the minimum number of stimuli of any dataset), then split each subset into halves and calculated the correlation matrix for each split-half. We then computed the Kendall's tau-b coefficient between the lower triangles of the two correlation matrices, as a measure of reliability.

To test whether each observed Kendall's tau was significantly different from chance, we used permutation tests. By permuting the trait labels and recalculating Kendall's tau for each permuted dataset, we created a sampling distribution of Kendall's tau values under the null, from which a p value can be derived.

To do this, after generating the random split-halves of data, we permuted the column names (trait labels) for one of the split-halves 10,000 times, then calculated the Kendall's tau between the correlation matrix for the permuted split-half and the correlation matrix for the other split-half, creating a sampling distribution of Kendall's tau values under the null (Figure 6). Finally, we compared the observed Kendall's tau to the null distribution to produce a p value. This allowed us to test how observed reliability compared to chance reliability for each dataset.

THE ROLE OF VALENCE IN TRAIT RATINGS OF UNFAMILIAR PEOPLE

Overall, trait ratings were more intercorrelated within the unfamiliar people domain, compared to the famous people domain. To further investigate the correlation structure for trait ratings of unfamiliar people, we built 13 linear models (one for each trait) that predicted trait ratings as a function of target valence (whether the unfamiliar person performed a positive or negative behavior). In addition, we conducted PCA on the 13 trait ratings, and examined component loadings as a function of target valence.

CORRELATION STRUCTURES AFTER REMOVING VALENCE INFORMATION

It appeared that a single feature, valence, was capturing most of the variance in trait judgments for unfamiliar people. To examine whether there was a reliable structure in trait ratings of unfamiliar people even after removing valence information, we divided the trait rating data for unfamiliar people into two subsets—targets who performed positive behaviors, and targets who performed negative behaviors—then tested for reliable structure within each valence subset. As a complementary analysis, we removed the first PC from the trait rating data, then tested for remaining reliable structure. To do this, we (1) projected the trait rating data onto PC space; (2) removed the first PC by zeroing out all values; and (3) rotated the data back to their original coordinates using the transpose of the PCA rotation matrix.

PREDICTING HOLISTIC SIMILARITY AFTER REMOVING VALENCE INFORMATION

Given that valence may be driving perceptions of similarity between pairs of unfamiliar people, we tested whether pairs of unfamiliar people that were concordant in valence (i.e., both positive/negative/neutral) were associated with greater holistic similarity ratings, compared to counter-valenced pairs of unfamiliar people.

Then, to test whether trait distances could still predict holistic similarity after removing valence information, we added concordance in valence (i.e., whether two targets were of the same valence, or counter-valenced) as a covariate to each single-trait model. As a complementary analysis, we tested how well trait

distances predicted holistic similarity between pairs of positive unfamiliar people (162 pairs) and between pairs of negative unfamiliar people (163 pairs).

RESULTS

TRAIT RATINGS OF UNFAMILIAR PEOPLE: COMPARING TWO DATASETS

We first assessed whether trait ratings of unfamiliar people with and without names and faces were comparable. We found that, for each of the 13 traits, there was a significant correlation between ratings of unnamed targets, and ratings of named targets (Figure S1 in the supplementary material). In addition, for each of the 13 traits, there was a significant correlation between trait distances calculated for pairs of unnamed targets, and trait distances calculated for pairs of named targets (see Figure S1 in the supplementary material). Furthermore, for each dataset, we generated a correlation matrix comprising Pearson's correlation coefficients for all pairwise combinations of the 13 trait ratings. These two correlation matrices were highly concordant with each other (Kendall's $\tau = 0.857$, $p < .0001$). These results suggest that further analyses conducted on these two datasets will be comparable.

PREDICTING HOLISTIC SIMILARITY: WITHIN THE SET OF UNFAMILIAR PEOPLE (WITHOUT NAMES AND FACES)

We found that for each of the 13 traits, pairwise trait distance significantly predicted pairwise holistic similarity. For instance, if two unfamiliar people were given similar openness ratings, these targets were also perceived to be similar overall (by a separate group of participants); if two targets were given dissimilar openness ratings, they were perceived to be dissimilar overall. See Table 1 for statistics for each model, and Figure 1a for a scatterplot of holistic similarity versus distance along openness.

In addition, a cumulative model containing all 13 trait distances significantly predicted holistic similarity, $F(13, 886) = 280.80$, $p < .0001$, coefficient of determination (CoD) = 0.800. See Table 2 for detailed statistics, and Figure 1b for a scatterplot.

PREDICTING HOLISTIC SIMILARITY: WITHIN THE SET OF UNFAMILIAR PEOPLE (WITH NAMES AND FACES)

We found that for each of the 13 traits, pairwise trait distance significantly predicted pairwise holistic similarity (Table S1 in the supplementary material). A cumulative model containing all 13 trait distances significantly predicted holistic similarity as well, $F(13, 886) = 267.70$, $p < .0001$, CoD = 0.792; see Table S2 in the supplementary material. Thus, adding names and faces to the unfamiliar targets did not produce qualitatively different results. It is important to note, however,

TABLE 1. Results From 13 Linear Models Predicting Holistic Similarity Between Pairs of Unfamiliar People, Using Pairwise Trait Distance

Trait	Theory	*b*	*SE*	*t*	*p*	Adjusted *p*
Openness	Big 5	0.721	0.023	31.223	1.70E-145	5.10E-145
Conscientiousness	Big 5	0.871	0.016	53.230	5.42E-280	6.50E-279
Extraversion	Big 5	0.078	0.033	2.354	0.019	0.019
Agreeableness	Big 5	0.880	0.016	55.451	3.20E-292	4.16E-291
Neuroticism	Big 5	0.842	0.018	46.714	1.41E-242	9.87E-242
Dominance	face perception	0.408	0.030	13.400	1.85E-37	3.70E-37
Trustworthiness	face perception	0.858	0.017	50.113	2.22E-262	2.00E-261
Warmth	stereotype content model	0.866	0.017	51.981	5.37E-273	5.37E-272
Competence	stereotype content model	0.844	0.018	47.218	1.50E-245	1.20E-244
Agency	mind perception	0.740	0.022	32.924	1.51E-156	7.55E-156
Experience	mind perception	0.735	0.023	32.475	1.24E-153	4.96E-153
Intelligence	n/a	0.822	0.019	43.228	1.25E-221	7.50E-221
Attractiveness	n/a	0.867	0.017	52.152	5.84E-274	6.42E-273

Note. *p* values were corrected using the Holm-Bonferroni method.

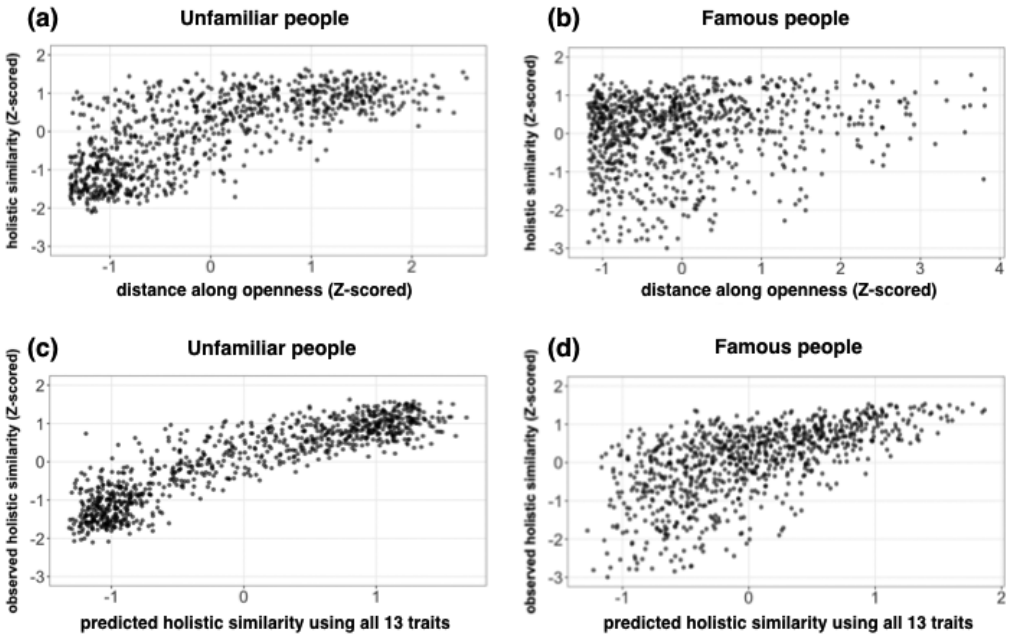

FIGURE 1. (a) Holistic similarity between pairs of unfamiliar people versus their distance along openness. Openness is being used as an illustrative example. (b) Holistic similarity between pairs of famous people versus their distance along openness. (c) Observed holistic similarity between pairs of unfamiliar people versus holistic similarity predicted by the cumulative model. (d) Observed holistic similarity between pairs of famous people versus holistic similarity predicted by the cumulative model.

TABLE 2. Results From a Cumulative Linear Model Predicting Holistic Similarity Between Pairs of Unfamiliar People, Using All 13 Pairwise Trait Distances

Variable	*b*	*SE*	*t*	*p*	Partial correlation
(Intercept)	0.000	0.015	0.000	1.000	
Openness	–0.009	0.033	–0.282	.778	–0.009
Conscientiousness	0.270	0.066	4.101	4.49E-05***	0.136
Extraversion	0.052	0.016	3.301	.001**	0.110
Agreeableness	0.235	0.099	2.361	.018*	0.079
Neuroticism	0.091	0.046	1.979	.048*	0.066
Dominance	0.101	0.018	5.670	1.93E-08***	0.187
Trustworthiness	0.167	0.053	3.150	.002**	0.105
Warmth	0.051	0.080	0.632	.527	0.021
Competence	0.129	0.061	2.111	.035*	0.071
Agency	–0.045	0.035	–1.299	.194	–0.044
Experience	0.027	0.036	0.729	.466	0.024
Intelligence	–0.050	0.049	–1.002	.317	–0.034
Attractiveness	0.008	0.079	0.101	.920	0.003

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

that we did not collect holistic similarity ratings for named targets; therefore, these models predicted holistic similarity between unnamed targets using trait distance between named targets. For this reason, in ensuing sections, we focus on discussing analyses of the unnamed target data; we note instances where these analyses were replicated on the named target data.

PREDICTING HOLISTIC SIMILARITY: WITHIN THE SET OF FAMOUS PEOPLE

For each of the 13 traits, pairwise trait distance significantly predicted pairwise holistic similarity (Table 3; Figure 1c). A cumulative model containing all 13 trait distances (Table 4; Figure 1d) significantly predicted holistic similarity, $F(13, 886) = 46.57$, $p < .0001$, CoD = 0.390.

PREDICTING HOLISTIC SIMILARITY: WITHIN THE SET OF FAMOUS PEOPLE, CONTROLLING FOR BIOGRAPHICAL INFORMATION

For each pair of famous people, we coded whether or not the targets shared the same gender, race, nationality, and industry (arts, athletics, business, media, politics, sciences). See Figure S10 in the supplementary material for visualizations of these pairwise biographical similarities.

TABLE 3. Results From 13 Linear Models Predicting Holistic Similarity Between Pairs of Famous People, Using Pairwise Trait Distance

Trait	Theory	*b*	*SE*	*t*	*p*	Adjusted *p*
Openness	Big 5	0.202	0.033	6.180	9.70E-10	2.91E-09
Conscientiousness	Big 5	0.355	0.031	11.367	4.53E-28	4.53E-27
Extraversion	Big 5	0.239	0.032	7.380	3.61E-13	1.81E-12
Agreeableness	Big 5	0.205	0.033	6.291	4.91E-10	1.96E-09
Neuroticism	Big 5	0.124	0.033	3.759	1.82E-04	1.82E-04
Dominance	face perception	0.446	0.030	14.917	4.05E-45	5.27E-44
Trustworthiness	face perception	0.243	0.032	7.492	1.62E-13	1.06E-12
Warmth	stereotype content model	0.154	0.033	4.671	3.46E-06	6.92E-06
Competence	stereotype content model	0.335	0.031	10.642	5.47E-25	4.92E-24
Agency	mind perception	0.243	0.032	7.502	1.51E-13	1.06E-12
Experience	mind perception	0.392	0.031	12.751	2.39E-34	2.63E-33
Intelligence	n/a	0.424	0.030	14.044	1.20E-40	1.44E-39
Attractiveness	n/a	0.317	0.032	10.010	1.97E-22	1.58E-21

Note. *p* values were corrected using the Holm-Bonferroni method.

We found that, for all traits other than neuroticism, pairwise trait distance significantly predicted pairwise holistic similarity, even after controlling for whether the two targets had the same gender, race, nationality, and industry (see Table S3 in the supplementary material for statistics). Thus, associations between holistic similarity and trait distance were largely robust to controlling for biographical similarities. As biographical information does not exist for the unfamiliar targets, we focus on discussing the performance of models that do not include biographical similarities as covariates.

PREDICTIVE PERFORMANCE: WITHIN-DOMAIN GENERALIZATION

Five-fold cross-validation was used to examine the within-domain predictive performance of models that predict holistic similarity using trait distance. Table 5 lists the cross-validated coefficient of determination (CoD), root mean squared error (RMSE), and mean absolute error (MAE) for each model in each domain.

We found that, for all traits other than dominance and extraversion, trait distance explained a greater proportion of variance in holistic similarity in the domain of unfamiliar people, than in the domain of famous people (see Figure 2a and 2b for radar plots of performance measures). In the domain of unfamiliar people, the top-performing traits in terms of predicting holistic similarity were: agreeableness, conscientiousness, attractiveness, warmth, and trustworthiness. In the domain of famous people, the top-performing traits were: dominance, intelligence,

TABLE 4. Results From a Cumulative Linear Model Predicting Holistic Similarity Between Pairs of Famous People, Using All 13 Pairwise Trait Distances

Variable	*b*	*SE*	*t*	*p*	Partial correlation
(Intercept)	0.000	0.026	0.000	1.000	
Openness	0.188	0.033	5.793	9.62E-09***	0.191
Conscientiousness	0.145	0.047	3.083	.002**	0.103
Extraversion	0.135	0.029	4.624	4.32E-06***	0.154
Agreeableness	−0.105	0.080	−1.312	.190	−0.044
Neuroticism	−0.055	0.039	−1.442	.150	−0.048
Dominance	0.330	0.031	10.571	1.11E-24***	0.335
Trustworthiness	0.100	0.057	1.747	.081	0.059
Warmth	0.029	0.058	0.491	.624	0.016
Competence	−0.329	0.068	−4.827	1.63E-06***	−0.160
Agency	−0.188	0.043	−4.415	1.13E-05***	−0.147
Experience	0.218	0.055	3.996	6.99E-05***	0.133
Intelligence	0.355	0.059	6.061	2.01E-09***	0.200
Attractiveness	0.180	0.028	6.549	9.81E-11***	0.215

Note. **$p < .01$. ***$p < .001$.

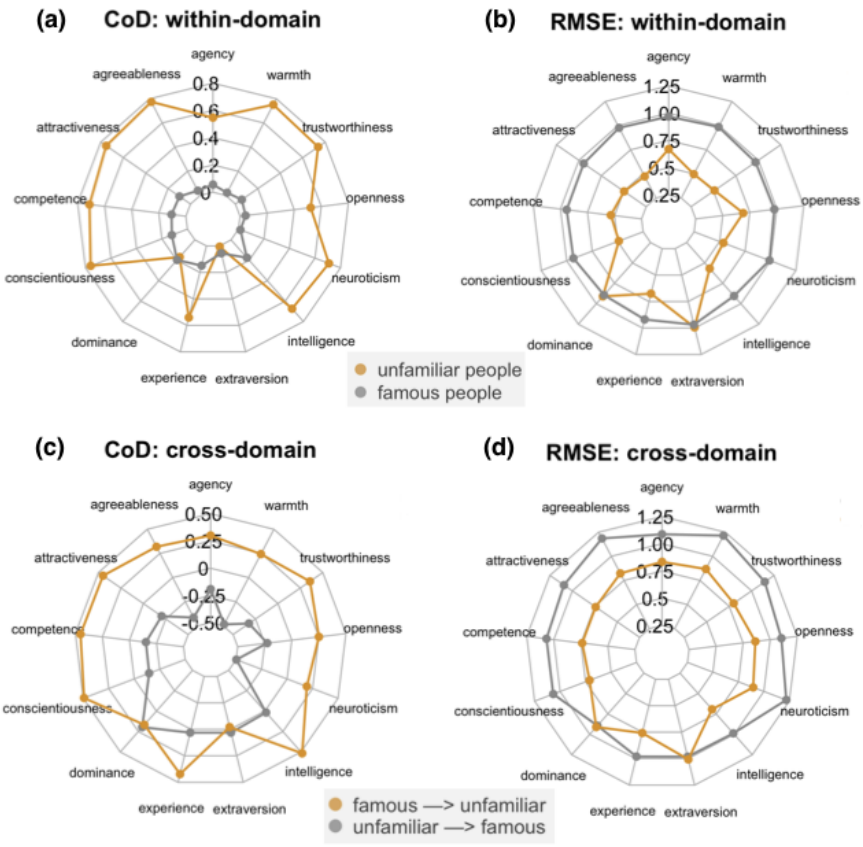

FIGURE 2. Performance measures for models predicting holistic similarity, visualized as radar plots. (a) CoD values by trait, within the domain of unfamiliar people and within the domain of famous people. (b) RMSE values by trait, within the domain of unfamiliar people and within the domain of famous people. (c) Cross-domain CoD values between observed and predicted holistic similarity values. (d) Cross-domain RMSE values between observed and predicted holistic similarity values.

TABLE 5. Within-Domain Predictive Performance of Models Predicting Pairwise Holistic Similarity Using Pairwise Trait Distance

Trait	CoD: unfamiliar people	CoD: famous people	RMSE: unfamiliar people	RMSE: famous people	MAE: unfamiliar people	MAE: famous people
Openness	0.522	0.040	0.689	0.977	0.558	0.778
Conscientiousness	0.759	0.122	0.488	0.934	0.385	0.734
Extraversion	0.005	0.056	0.995	0.969	0.879	0.770
Agreeableness	0.775	0.039	0.473	0.977	0.372	0.782
Neuroticism	0.709	0.015	0.538	0.990	0.433	0.788
Dominance	0.168	0.197	0.909	0.894	0.777	0.694
Trustworthiness	0.737	0.057	0.511	0.968	0.409	0.767
Warmth	0.751	0.022	0.497	0.986	0.392	0.786
Competence	0.713	0.106	0.534	0.942	0.424	0.743
Agency	0.547	0.056	0.670	0.969	0.539	0.770
Experience	0.541	0.149	0.676	0.920	0.552	0.719
Intelligence	0.675	0.175	0.568	0.905	0.454	0.704
Attractiveness	0.752	0.098	0.496	0.947	0.393	0.745
All 13	0.800	0.390	0.446	0.778	0.352	0.602

Note. Five-fold cross-validation was used to calculate performance measures. The bottom row reports performance for the cumulative model. CoD: coefficient of determination; RMSE: root mean squared error; MAE: mean absolute error.

experience, conscientiousness, and competence. In addition, the cumulative model (containing all 13 trait distances) explained a greater proportion of variance in holistic similarity in the domain of unfamiliar people, than in the domain of famous people.

These results were largely replicated when names and faces were added to the unfamiliar targets: For all traits other than dominance, trait distance explained a greater proportion of variance in holistic similarity in the domain of unfamiliar people, and the cumulative model explained a greater proportion of variance in holistic similarity in the domain of unfamiliar people (Table S5; Figure S4 in the supplementary material).

PARTIAL CORRELATIONS IN CUMULATIVE MODELS

Figure 3 plots, for each domain, partial correlations between each trait distance and holistic similarity, controlling for the other 12 trait distances. These partial effects within cumulative models provide one way to evaluate the relative importance of different traits for perceived similarity. In the domain of unfamiliar people, the traits with the largest partial effects were: dominance, conscientiousness, extraversion, trustworthiness, and agreeableness. In the domain of famous people, they were: dominance, attractiveness, intelligence, openness, and extraversion.

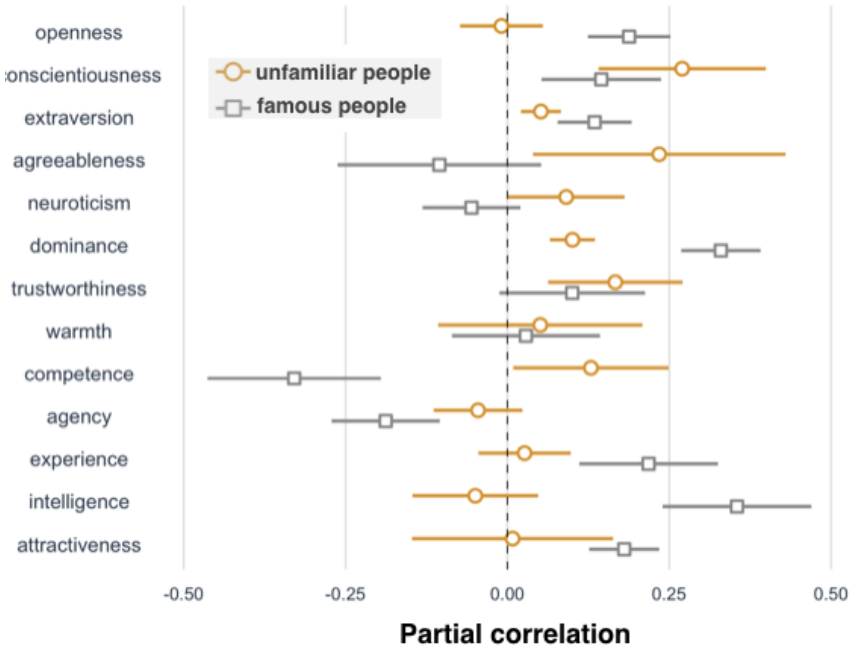

FIGURE 3. Partial correlations between each trait distance and holistic similarity, controlling for the other 12 trait distances, in the domain of unfamiliar people and in the domain of famous people.

When names and faces were added to the unfamiliar targets, three of the partial effects changed in significance (Table S2 and Figure S2 in the supplementary material; trustworthiness became nonsignificant, while warmth and agency became significant). When biographical similarity was controlled for among famous targets, four of the partial effects changed in significance (Table S4 and Figure S3 in the supplementary material; conscientiousness became nonsignificant, while agreeableness, neuroticism, and trustworthiness became significant).

The reported partial effects indicate the unique contributions of each trait to holistic similarity, over and above the other traits; however, these partial effects depend on the particular set of 13 traits that were tested. Thus, when evaluating the relative importance of different traits for perceived similarity, we will focus on the predictive performance of single-trait models.

PREDICTIVE PERFORMANCE: CROSS-DOMAIN GENERALIZATION

Five-fold cross-validation was used to examine the cross-domain predictive performance of models that predict holistic similarity using trait distance. See Table 6 for cross-validated performance measures for each model in each domain, and Figure 4 for plots of predicted versus observed holistic similarity.

TABLE 6. Cross-Domain Predictive Performance of Models Predicting Pairwise Holistic Similarity Using Trait Distance

Trait	CoD: unfamiliar people → famous people	CoD: famous people → unfamiliar people	RMSE: unfamiliar people → famous people	RMSE: famous people → unfamiliar people	MAE: unfamiliar people→ famous people	MAE: famous people → unfamiliar people
Openness	−0.227	0.251	1.104	0.863	0.900	0.752
Conscientiousness	−0.147	0.492	1.067	0.711	0.869	0.604
Extraversion	0.031	−0.021	0.982	1.008	0.778	0.883
Agreeableness	−0.414	0.319	1.185	0.823	0.980	0.715
Neuroticism	−0.497	0.194	1.220	0.895	1.003	0.784
Dominance	0.196	0.167	0.894	0.910	0.695	0.774
Trustworthiness	−0.323	0.357	1.146	0.799	0.918	0.693
Warmth	−0.485	0.243	1.215	0.868	0.999	0.760
Competence	−0.152	0.452	1.068	0.738	0.856	0.629
Agency	−0.192	0.300	1.088	0.835	0.885	0.727
Experience	0.031	0.422	0.980	0.758	0.782	0.649
Intelligence	0.017	0.517	0.987	0.693	0.784	0.584
Attractiveness	−0.204	0.449	1.093	0.740	0.888	0.635
All 13	−0.034	0.620	1.013	0.615	0.823	0.495

Note. Five-fold cross-validation was used to calculate performance measures. A negative coefficient of determination indicates poorer prediction than the mean value. The bottom row reports performance for the cumulative model. CoD: coefficient of determination; RMSE: root mean squared error; MAE: mean absolute error.

We found that, for all traits other than dominance and extraversion, the mapping between trait distance and holistic similarity generalized better from a training set of famous people to a testing set of unfamiliar people, than from a training set of unfamiliar people to a testing set of famous people (Figure 2c and 2d; Table 6). In addition, the cumulative (13-trait) model generalized better from a training set of famous people to a testing set of unfamiliar people, than vice versa (Table 6).

The above results were largely replicated when names and faces were added to the unfamiliar targets: For all traits other than dominance, models generalized better from the famous people data to the unfamiliar people data, and the cumulative model generalized better from the famous people data to the unfamiliar people data (Table S6; Figure S4 in the supplementary material).

ACCOUNTING FOR DIFFERENCES IN GENERALIZATION

The asymmetry in cross-domain generalization was pronounced. Thirteen of the models trained on the unfamiliar people data (all models except the dominance model) produced negative or near-zero coefficients of determination when predicting holistic similarity for famous people (range of CoDs: −0.497 to 0.031; Table 6, column 1). That is, most models exhibited poorer prediction performance than a model that just predicts the mean value for holistic similarity. In contrast, 13 of

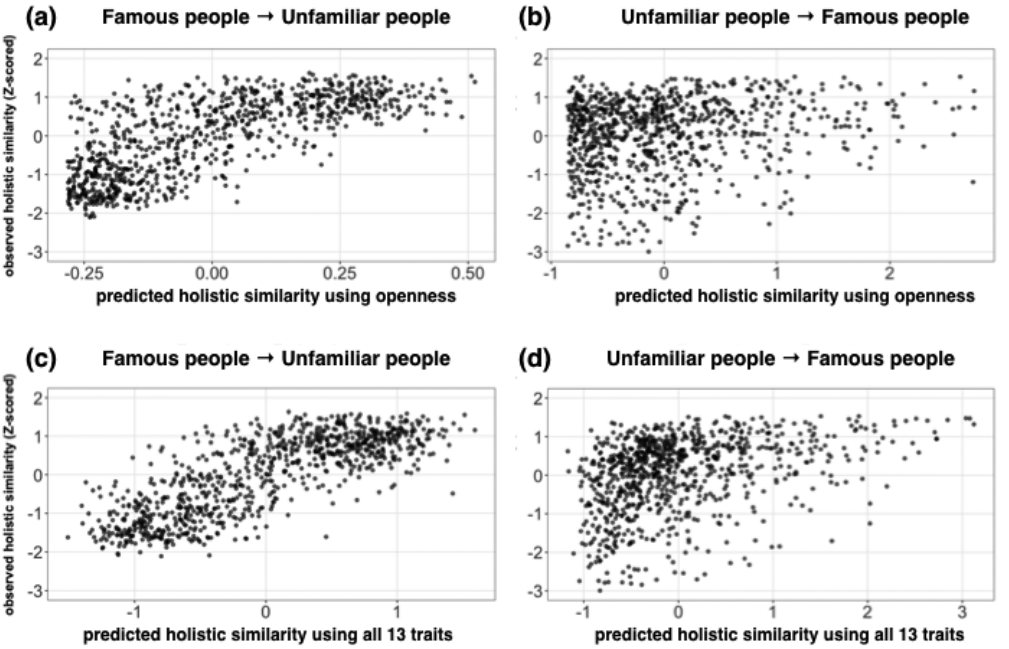

FIGURE 4. (a) Observed holistic similarity between pairs of unfamiliar people versus holistic similarity predicted by a single-trait model trained on famous people. (b) Observed holistic similarity between pairs of famous people versus holistic similarity predicted by a single-trait model trained on unfamiliar people. (c) Observed holistic similarity between pairs of unfamiliar people versus holistic similarity predicted by a 13-trait model trained on famous people. (d) Observed holistic similarity between pairs of famous people versus holistic similarity predicted by a 13-trait model trained on unfamiliar people.

the models trained on the famous people data (all models except the extraversion model) produced positive coefficients of determination when predicting holistic similarity for unfamiliar people—here, model predictions performed better than a model that just predicts the mean value (range of CoDs: 0.167 to 0.620; Table 6, column 2). This asymmetry in prediction accuracy is consistent with the presence of an asymmetry in the dimensionality of the two representational spaces.

If we had observed that prediction accuracy was at or below chance for both directions of generalization (unfamiliar-to-famous and famous-to-unfamiliar), we would not have been able to infer that the two representational spaces had dimensions in common. However, we instead observed that nearly all models that were trained on the famous people data were able to predict holistic similarity between unfamiliar people with above-chance accuracy. For this reason, we hypothesize that the representational space of famous people includes dimensions from the representational space of unfamiliar people, *as well as* other dimensions—that is, famous people may be represented in a higher-dimensional space than unfamiliar people are. This could account for the asymmetry where generalization from famous people to unfamiliar people is more successful than generalization from unfamiliar people to famous people: Inferring a higher-dimensional space from a

lower-dimensional (perhaps one-dimensional) space may pose a more challenging prediction task compared to the opposite direction.

Our proposed hypothesis—that famous people are represented in a higher-dimensional space—can account for low prediction accuracy during generalization from unfamiliar people to famous people, but it does not explain why the coefficients of determination for many of these models are negative, rather than positive and near zero. To investigate this phenomenon more closely, we examined the distribution of holistic similarity ratings for each domain. We found that for the domain of unfamiliar people, the distribution of holistic similarity ratings had a skew of −0.325 (indicating a slightly long left tail) and a kurtosis of 1.760 (indicating thinner tails and a broader peak). For the domain of famous people, the distribution of holistic similarity ratings had a skew of −0.982 (indicating a moderately long left tail) and a kurtosis of 3.544 (indicating fatter tails and a sharper peak). Thus, both distributions showed some non-normality, which may have contributed to negative coefficients of determination. However, it is important to note that non-normality should contribute to below-chance performance for both directions of generalization (unfamiliar-to-famous and famous-to-unfamiliar). Therefore, non-normality alone is not sufficient to explain (1) the marked asymmetry observed in generalization performance, or (2) the large coefficients of determination observed when generalizing from famous to unfamiliar people. Our proposed hypothesis, of a difference in representational complexity, is further explored below.

CORRELATION STRUCTURES

We next examined the collinearity of trait ratings in each domain. As seen in Figure 5a and 5b, the Pearson's correlations between trait ratings of unfamiliar people were more extreme than the correlations between trait ratings of famous people. A chi-squared test of the two correlation matrices revealed that they significantly differed, $\chi^2(78) = 4881.09$, $p < .0001$. PCA revealed that, in each domain, the first principal component accounted for a majority of the variance in trait ratings (see Figure 5c and 5d for scree plots; see Tables S13 and S14 in the supplementary materials for component loadings). The first principal component (PC) accounted for a greater proportion of variance in the unfamiliar people domain (0.835) compared to the famous people domain (0.603).

RELIABILITY OF CORRELATION STRUCTURES

Next, we examined whether the degree of intercorrelatedness between trait ratings is a robust feature of each domain, rather than being variable across different subsets of data. We found that, for the unfamiliar people domain and the famous people domain, the correlation matrices for random split-halves of data were significantly correlated (unfamiliar people: Kendall's $\tau = 0.849$, permutation $p < .0001$; famous people: Kendall's $\tau = 0.782$, permutation $p < .0001$; Figure 6a and 6b). These results indicate that the degree of intercorrelatedness between traits is a robust feature in each domain. The above results were replicated when names

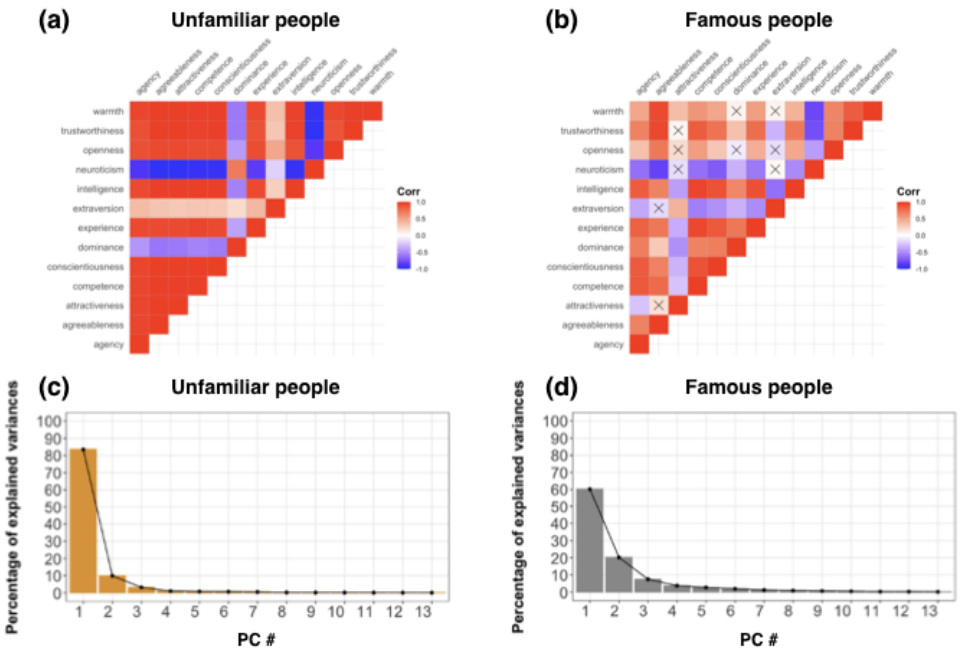

FIGURE 5. Top panels: Pearson's correlations for all pairwise combinations of the 13 trait ratings in the domain of (a) unfamiliar people and (b) famous people. Xs refer to non-significant correlations. Bottom panels: Scree plots displaying proportion of total variance explained by each principal component in the domain of (c) unfamiliar people and (d) famous people.

and faces were added to the unfamiliar targets (Figure S6; Figure S7 in the supplementary material).

THE ROLE OF VALENCE IN TRAIT RATINGS OF UNFAMILIAR PEOPLE

Taking a closer look at the correlation structure for trait ratings of unfamiliar people (Figure 5a), we found that 11 of the traits were positively correlated with each other, but largely anticorrelated with neuroticism and dominance (dominance and extraversion, however, were positively correlated). In addition, when trait ratings were regressed onto target valence (whether the target performed a positive or negative behavior), we found that negative targets (vs. positive targets) were rated significantly higher on neuroticism and dominance (neuroticism: $b = -1.915$, $SE = 0.035$, $t = -54.62$, $p < .0001$; dominance: $b = -1.353$, $SE = 0.091$, $t = -14.81$, $p < .0001$), whereas positive targets were rated significantly higher on all other traits (see Table S15 in the supplementary material for statistics).

Furthermore, we found that the first PC loaded positively onto neuroticism and dominance, but negatively onto all other traits (Table S13 in the supplementary material), and the first PC scores were positive for negative targets but negative for positive targets (Figure S5). In all, this pattern of results indicates that a single

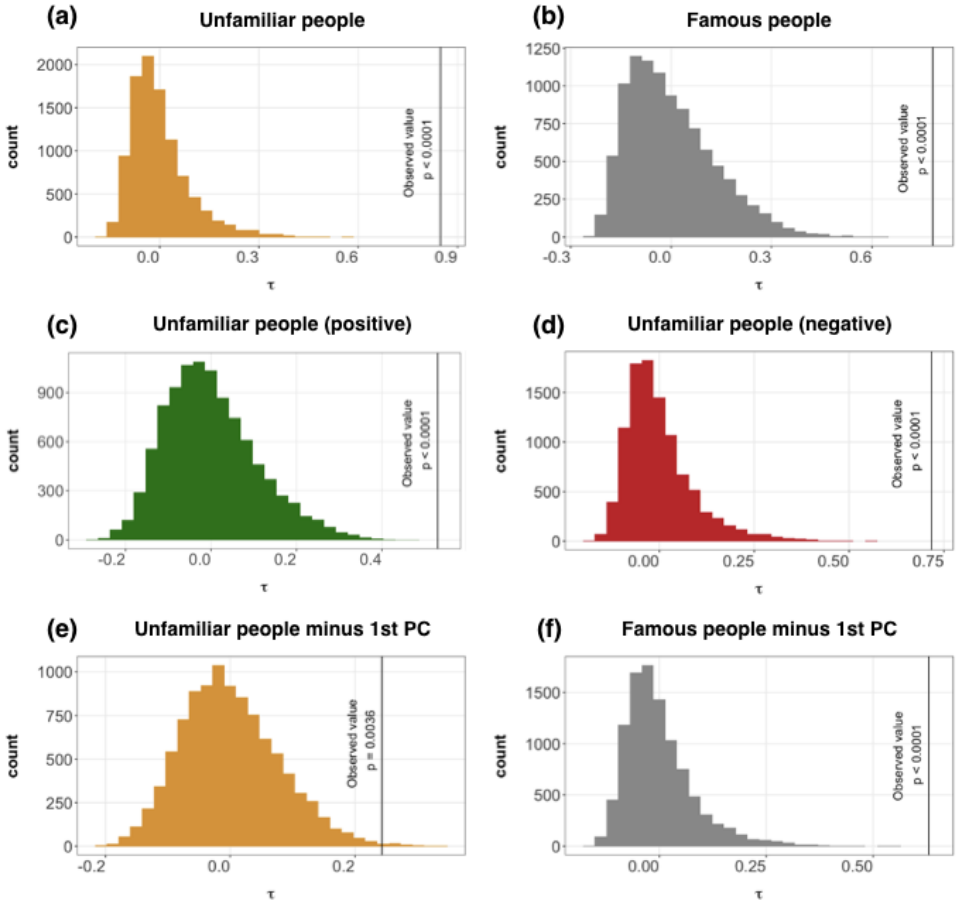

FIGURE 6. Kendall's tau distributions for permuted data in the domain of (a) unfamiliar people, (b) famous people, (c) unfamiliar people who performed positive behaviors, (d) unfamiliar people who performed negative behaviors, (e) unfamiliar people after the first PC was removed, (f) famous people after the first PC was removed.

underlying feature, valence, is capturing most of the variation in trait judgments in the domain of unfamiliar people.

CORRELATION STRUCTURES AFTER REMOVING VALENCE INFORMATION

Overall, the 13 trait ratings were more intercorrelated within the unfamiliar people domain, compared to the famous people domain, indicating that the unfamiliar targets may reside in a lower-dimensional (perhaps one-dimensional) representational space. This asymmetry might explain why it is (1) more accurate to use trait distances to predict holistic similarity between unfamiliar people than between famous people, and (2) more accurate to use models trained on famous people

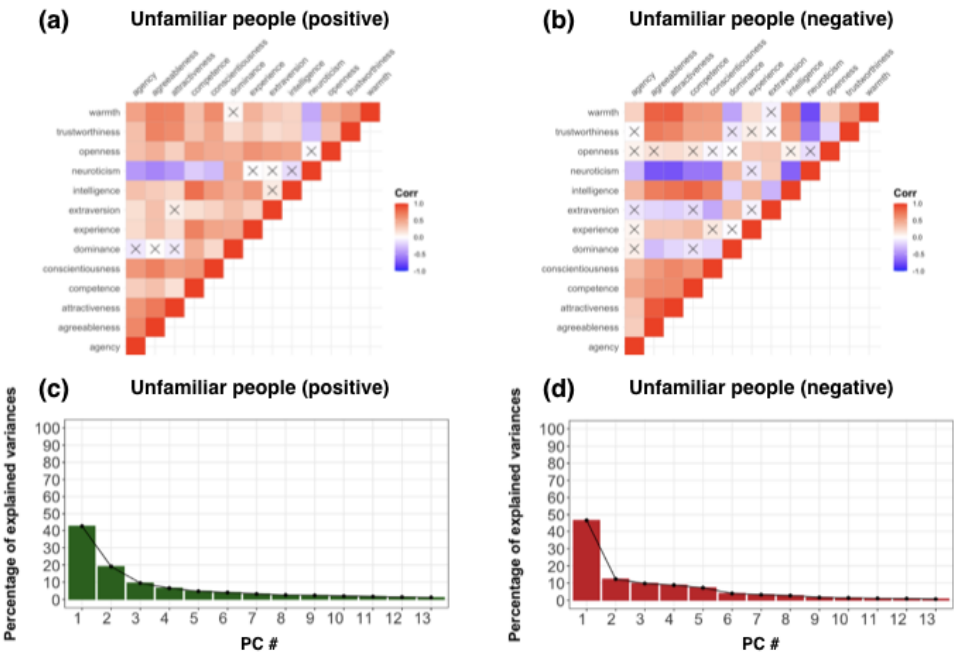

FIGURE 7. Top panels: Pearson's correlations for all pairwise combinations of the 13 trait ratings for (a) unfamiliar people who performed positive behaviors and (b) unfamiliar people who performed negative behaviors. Xs refer to non-significant correlations. Bottom panels: Scree plots displaying proportion of total variance explained by each principal component in the domain of (c) unfamiliar people who performed positive behaviors and (d) unfamiliar people who performed negative behaviors.

to predict holistic similarity between unfamiliar people than vice versa. To test whether the representational space for unfamiliar people is one-dimensional, we examined if there is remaining reliable structure in trait ratings of unfamiliar people even after removing the feature that seems to account for most of the variance—valence.

Separate correlation matrices were generated for the subset of positive unfamiliar people, and the subset of negative unfamiliar people. We found that there was reduced intercorrelation between trait ratings of unfamiliar people of the same valence (Figure 7a and 7b). The following comparisons between correlation matrices were significant: between all unfamiliar people and positive unfamiliar people, $\chi^2(78) = 9{,}917.78$, $p < .0001$; between all unfamiliar people and negative unfamiliar people, $\chi^2(78) = 8{,}992.06$, $p < .0001$; and between positive unfamiliar people and negative unfamiliar people, $\chi^2(78) = 640.47$, $p < .0001$. As illustrated by the scree plots (Figure 7c and 7d), the first PC for each valence subset explained less than half of all variance (positive unfamiliar people: 42.6%; negative unfamiliar people: 46.5%), whereas in the set of all unfamiliar people, the first PC had explained more than 80% of all variance.

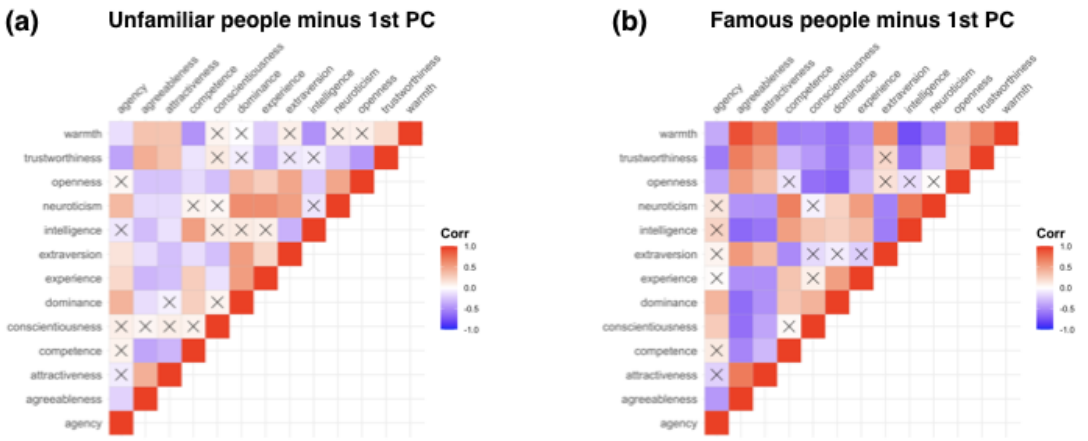

FIGURE 8. Pearson's correlations for all pairwise combinations of the 13 trait ratings. (a) Ratings for the domain of unfamiliar people, after removing the first PC and (b) ratings for the domain of famous people, after removing the first PC.

We examined the reliability of the correlation structure for each valence subset. The correlation matrices for random split-halves of data were significantly correlated for each valence subset (positive unfamiliar people: Kendall's $\tau = 0.530$, permutation $p < .0001$; negative unfamiliar people: Kendall's $\tau = 0.717$, permutation $p < .0001$; Figure 6b and 6c). This indicates that the correlation structure of each valence subset is robust. These results were replicated when names and faces were added to the unfamiliar targets (Figure S7; Figure S8 in the supplementary material).

However, there might still be variance along the valence axis within the subset of positive unfamiliar people, and within the subset of negative unfamiliar people—some positive unfamiliar people are more positive than others, and some negative unfamiliar people are more negative than others. Thus, as a stricter test, we removed the first PC (Tables S13 and S14 in the supplementary material) from the trait rating data for each domain, then tested for remaining reliable structure (Figure 8).

After removing the first PC, the correlation matrix for unfamiliar people was significantly different from the original, $\chi^2(78) = 30{,}760.81$, $p < .0001$, and the correlation matrix for famous people was significantly different from the original, $\chi^2(78) = 1{,}759.05$, $p < .0001$. For both domains, removing the first PC still resulted in reliable correlation structures (unfamiliar people: Kendall's $\tau = 0.243$, permutation $p = .0036$; famous people: Kendall's $\tau = 0.630$, permutation $p < .0001$; Figure 6e and 6f). Thus, for both the unfamiliar people domain and the famous people domain, the trait rating data exhibited a reliable structure even after removing the first PC. Overall, we found that a reliable higher-dimensional structure exists for both unfamiliar targets and famous targets; valence was not the *only* feature driving trait judgments. These results were replicated when names and faces were added to the unfamiliar targets (Figure S7; Figure S9 in the supplementary material).

PREDICTING HOLISTIC SIMILARITY AFTER REMOVING VALENCE INFORMATION

We next tested whether valence was driving perceptions of similarity between unfamiliar people. We found that pairs of unfamiliar people that performed behaviors of the same valence were rated as more similar overall, compared to pairs that performed counter-valenced behaviors, $b = –1.754$, $F(1.898) = 2.316$, $p < .0001$, $R^2 = 0.721$. Again, however, valence was not the *only* feature that mattered: When concordance in valence was added as a covariate to each single-trait model, 12 of the trait distances still significantly predicted holistic similarity (Table S7 in the supplementary material).

As a complementary analysis, we tested how well trait distances predicted holistic similarity within each valence subset. For pairs of positive unfamiliar people, four of the trait distances significantly predicted holistic similarity (Table S8 in the supplementary material), and the 13-trait model significantly predicted holistic similarity, $F(13,148) = 2.559$, $p = .0033$, $R^2 = 0.184$. For pairs of negative unfamiliar people, six of the trait distances significantly predicted holistic similarity (Table S9 in the supplementary material), and the 13-trait model significantly predicted holistic similarity, $F(13,149) = 3.105$, $p = .0004$, $R^2 = 0.213$. Thus, when positive and negative unfamiliar people were separated, trait distances performed worse in predicting holistic similarity; however, some traits still significantly predicted holistic similarity. The above results were replicated when names and faces were added to the unfamiliar targets (Tables S10–S12 in the supplementary material).

DISCUSSION

In the current study, we examined how 13 traits from major theories of person perception (Thornton & Mitchell, 2018) contributed to overall representations of famous people and to overall representations of unfamiliar people, by probing how well each trait predicted perceived similarity between pairs of targets. This approach allowed us to examine the importance of different traits in determining perceivers' overall representations of people, and whether the representational structure depended on inference context.

Previous research on the structure of trait representations has relied on reducing the dimensionality of evaluations of targets along particular traits. However, such evaluations do not contain information about the relative importance of a particular trait for perceivers' overall representations of people; the dimensions that explain the most variance across behavioral judgments or neural responses may not necessarily contribute the most to overall representations. The current study deviates from previous research in that we used a perceived similarity approach to gauge the importance of different traits for overall representations of people. This method revealed that (1) 13 traits from extant models of person knowledge could each predict perceived similarity between pairs of targets; (2) the traits that best predicted holistic similarity between unfamiliar people were different from

the ones that best predicted holistic similarity between famous people; and (3) trait ratings were more intercorrelated for unfamiliar people than for famous people, suggesting that the representational structure of first impressions is largely driven by one feature, valence. However, further analyses showed that for trait ratings of both unfamiliar and famous targets, a reliable structure was present even after removing the first principal component, indicating a higher-dimensional structure for first impressions as well.

CONTRIBUTIONS OF TRAITS TO OVERALL REPRESENTATIONS

In the domain of unfamiliar people, we found that distance along each of 13 traits, individually and together, successfully predicted pairwise holistic similarity ratings; adding names and faces to the unfamiliar targets did not qualitatively change these results. In the domain of famous people, we also found that distance along each of 13 traits, individually and together, successfully predicted pairwise holistic similarity ratings; these associations were largely robust to controlling for pairwise biographical similarities.

The significant associations between pairwise trait distance and pairwise holistic similarity indicate that the 13 tested traits contribute to perceivers' overall representations of people, both when thinking about unfamiliar people, and when thinking about famous people. Importantly, the traits did not all perform equally well in predicting holistic similarity. Differences in predictive performance allow us to make inferences about the importance of particular traits for overall representations of people.

When examining the performance of single-trait models in the domain of unfamiliar people, the top-performing traits were: agreeableness, conscientiousness, attractiveness, warmth, and trustworthiness. These traits cut across multiple theories of person perception: the Big 5 (agreeableness and conscientiousness), the stereotype content model (warmth), and the model of face perception (trustworthiness). The top-performing traits in the domain of famous people were: dominance, intelligence, experience, conscientiousness, and competence. These traits again cut across multiple theories: the model of face perception (dominance), the model of mind perception (experience), the Big 5 (conscientiousness), and the stereotype content model (competence). Thus, the traits that best predicted holistic similarity between unfamiliar people were largely different from the ones that best predicted holistic similarity between famous people; in addition, within each domain, there was some conceptual overlap between the top-performing traits.

Previous research across different subfields of psychology has consistently shown that two fundamental dimensions seem to underlie social evaluations: communion (captured by traits that relate to morality and sociability, such as trustworthiness and warmth) and agency (captured by traits that relate to ability and assertiveness, such as competence and dominance; Abele & Wojciszke, 2014; Oliveira et al., 2020). These Big Two dimensions are thought to have functional significance, as communion-related traits describe whether an individual has good or bad intentions, and whether they can garner social support for their intentions,

while agency-related traits describe whether an individual can carry out their intentions, and how much power they have over others; assessments along these dimensions can carry consequences for motivations and behaviors toward individuals and groups (Abele & Wojciszke, 2014; Cuddy et al., 2008; Fiske et al., 2002, 2007; Landy et al., 2016; Oliveira et al., 2020).

It is of note that the traits that contribute most to holistic similarity for the studied set of famous people (dominance, intelligence, experience, conscientiousness, and competence) mostly fall under the "agency" umbrella of traits, while the traits that contribute most to holistic similarity for the studied set of unfamiliar people (agreeableness, conscientiousness, attractiveness, warmth, and trustworthiness) mostly fall under the "communion" umbrella of traits. Prior work probing the relationship between valence and the Big Two has shown that communion-related traits exhibit greater overlap with valence when compared to agency-related traits (Abele & Wojciszke, 2014); it may be that communion-related traits best predicted holistic similarity between unfamiliar targets in the current study, because valence was a key feature organizing trait judgments of unfamiliar targets. Below we return to the idea that members of the two studied domains are more differentiable along one umbrella dimension than the other.

ASYMMETRY IN GENERALIZATION PERFORMANCE

Trait distances, both together and individually, better predicted holistic similarity between unfamiliar people than between famous people. In addition, the mapping from trait distance to holistic similarity generalized better from a training set of famous people to a testing set of unfamiliar people, than vice versa. Notably, most of the models trained on the unfamiliar people data produced negative coefficients of determination when predicting holistic similarity for famous people (indicative of below-chance accuracy); in contrast, most of the models trained on the famous people data produced positive coefficients of determination when predicting holistic similarity for unfamiliar people.

As holistic similarity ratings in both domains exhibited moderate non-normality, non-normality alone is not sufficient to explain (1) the marked asymmetry observed in generalization performance, or (2) the above-chance accuracy observed when generalizing from famous to unfamiliar people. Our proposed explanation for poor generalization from the unfamiliar people domain to the famous people domain is that the unfamiliar targets are represented in a lower-dimensional space than the famous targets. In line with this, PCA of trait ratings in each domain revealed that the first PC accounts for a greater proportion of variance in the unfamiliar people domain, compared to the famous people domain. Furthermore, we found that trait ratings of unfamiliar people were more correlated with each other than trait ratings of familiar people were. These correlation structures were robust across random splits of data for both domains. This indicates that the greater intercorrelatedness between trait ratings of unfamiliar people was not just due to trait ratings of famous people being noisier; rather, the correlation structure of each domain was reliable.

DIMENSIONALITY OF EACH DOMAIN

Given the high intercorrelatedness between trait ratings of unfamiliar people, we also tested the more specific hypothesis that one feature, valence, was driving trait ratings similarity judgments in the domain of unfamiliar people. We found that concordance in valence between pairs of unfamiliar people explained 72% of the variance in pairwise holistic similarity ratings. That is, whether two unfamiliar targets performed behaviors of the same valence successfully predicted the holistic similarity for that pair. The relative ease of classifying the unfamiliar targets as good or bad is likely why the unfamiliar people data exhibited (1) greater correlations among trait ratings, and (2) a stronger relationship between trait distance and holistic similarity—unfamiliar people of the same valence likely received more similar trait ratings and higher holistic similarity ratings than unfamiliar people of the opposite valence. This also explains why communion-related traits (agreeableness, warmth, and trustworthiness) performed the best at predicting holistic similarity between unfamiliar people.

However, several additional analyses revealed that valence is not the *only* feature that matters for representations of unfamiliar people. First, we split the unfamiliar targets into positive and negative, removing the dominant organizational feature. We found that trait ratings were less correlated with each other in each valence subset, compared to the complete set of all unfamiliar people, but the correlation structure of each valence subset was still reliable. Second, we removed valence information in a different way, by removing the first PC from the trait ratings of unfamiliar people, and the resultant correlation structure was also reliable. Third, we found that, even when concordance in valence was added as a covariate to single-trait models predicting holistic similarity, 12 traits still significantly predicted holistic similarity. These results suggest that a higher-dimensional representational structure exists for the unfamiliar targets, but it occurs on top of a lower-dimensional structure organized along a positive–negative axis; the content of this higher-dimensional structure is an important open question.

While the higher-dimensional structure for unfamiliar people explains less variance in holistic similarity, we found that this structure is still reliable, and it may still play a key role in social judgments and predictions, such as in everyday contexts where people's behaviors may not be as clearly valenced as the positive and negative behaviors presented in this study. We hypothesize that even when a small number of dimensions can account for most of the variance in trait ratings, a much larger number of dimensions might still be reliable and crucial for accounting for human social judgments, even if the proportion of total variance in judgments they explain is small.

STIMULUS DEPENDENCE

We now turn to a key observation that needs to be taken into account when interpreting the results. Differences between the two domains are likely shaped by the sets of stimuli tested (e.g., see Lin et al., 2019, for evidence that surveying a larger

number of trait words than is typical yields a novel set of four dimensions that best explain trait judgments of faces). The unfamiliar people stimuli were designed to be highly valenced (largely positive or negative; Kim et al., 2021), and the famous people stimuli were designed to be a maximally varied collection of famous people (Thornton & Mitchell, 2018). Due to these differences in stimulus selection, the observed differences in terms of which traits best predicted holistic similarity should be interpreted with caution.

We note that there were some additional sources of variation among the unfamiliar people stimuli (as measured in Kim et al., 2021): The behaviors that were performed by targets varied in emotional arousal and perceived frequency. The emotional intensity of each behavior was rated on a scale from 1 to 7 (N = 30 participants/behavior), and the perceived frequency of each behavior was rated on a scale from 1 to 100 (N = 30 participants/behavior). The set of 300 behaviors displayed variance along both features (emotional arousal: M = 3.81, SD = 0.94, range = 1.25–6.28; perceived frequency: M = 24.08, SD = 19.04, range = 1–98.82). In addition, the subset of positive behaviors and the subset of negative behaviors did not significantly differ along these features ($p > .10$)—variation was distributed across valence, meaning that the sample of unfamiliar people in the current study was not necessarily predetermined to be one-dimensional. Thus, the unfamiliar people stimuli can provide us with limited but still useful insight into the representational structure of first impressions.

It is likely that, if a more varied set of behaviors were associated with the unfamiliar targets, the valence axis would have been less salient to perceivers, and we would have found a less robust relationship between trait judgments of unfamiliar people and holistic similarity judgments. However, it should be noted that the highly valenced nature of the unfamiliar people stimuli in the current study made it harder, rather than easier, to identify higher-dimensional structure beyond valence. Our results show that even in a set of stimuli that predominantly vary along the valence axis, higher-dimensional information inferred from behaviors was sufficiently strong to display reliable structure. That is, the finding that the representational structure is higher-dimensional for famous people versus unfamiliar people is less likely to be stimulus-bound. An important direction for future work is to utilize a more varied set of unfamiliar targets and to more rigorously examine both which traits are important for representations of unfamiliar people and how representations differ across contexts.

CONCLUSION

In this study, we used a perceived similarity approach to gauge the importance of different traits for overall representations of famous people and overall representations of unfamiliar people for whom one behavior is known. We found that (1) 13 traits from extant models of person knowledge could predict perceived similarity between pairs of targets; (2) the traits that best predicted holistic similarity partially depended on inference context (unfamiliar people vs. famous people); and (3) trait ratings were more intercorrelated for unfamiliar targets than

for famous targets, but reliable higher-dimensional structure was present even for first impressions. These findings highlight a new way to probe perceivers' overall representations of people, and shed light on how trait representations are affected by inference context.

REFERENCES

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology*, *50*, 195–255.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i–171. https://doi.org/10.1037/h0093360

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–903.

Gerstenberg, T., Ullman, T. D., Nagel, J., Kleiman-Weiner, M., Lagnado, D. A., & Tenenbaum, J. B. (2018). Lucky or clever? From expectations to responsibility judgments. *Cognition*, *177*, 122–141.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037//0022-3514.59.6.1216

Gorno-Tempini, M. L., & Price, C. J. (2001). Identification of famous faces and buildings: A functional neuroimaging study of semantically unique items. *Brain*, *124*(10), 2087–2097.

Grabowski, T. J., Damasio, H., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Human Brain Mapping*, *13*(4), 199–212.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Kim, M. J., Mende-Siedlecki, P., Anzellotti, S., & Young, L. (2021). Theory of mind following the violation of strong and weak prior beliefs. *Cerebral Cortex*, *31*(2), 884–898.

Kryven, M., Ullman, T., Cowan, W., & Tenenbaum, J. (2016, August 10-13). *Outcome or strategy? A Bayesian model of intelligence attribution* [Paper presentation]. CogSci 2016, Philadelphia, Pennsylvania.

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin*, *42*(9), 1272–1290.

Lin, C., Keles, U., & Adolphs, R. (2019, October 2). *Four dimensions characterize comprehensive trait judgments of faces*. PsyArXiv. https://doi.org/10.31234/osf.io/87nex

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces* [CD ROM]. Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, Stockholm, Sweden.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037//0022-3514.52.1.81

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, *8*(6), 623–631.

Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2020). Good to bad or bad to bad? What is the relationship between valence and the trait

content of the Big Two? *European Journal of Social Psychology*, *50*(2), 463–483.

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

Pagan, M., Simoncelli, E. P., & Rust, N. C. (2016). Neural quadratic discriminant analysis: Nonlinear decoding with v1-like computation. *Neural Computation*, *28*(11), 2291–2319.

Park, I. M., Meister, M. L., Huk, A. C., & Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*, *17*(10), 1395–1403.

R Core Team. (2013). *R: A language and environment for statistical computing.*

Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, *26*(3), 179–195.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212.

Thornton, M. A., & Mitchell, J. P. (2018). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex*, *28*(10), 3505–3520.

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The brain represents people as the mental states they habitually experience. *Nature Communications*, *10*(1), 2291. https://doi.org/10.1038/s41467-019-10309-7

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive Science*, *42*(3), 850–884.