The conceptual structure of face impressions

Ryan M. Stolier[a,1], Eric Hehman[b], Matthias D. Keller[c], Mirella Walker[c], and Jonathan B. Freeman[a,d,1]

[a]Department of Psychology, New York University, New York, NY 10003; [b]Department of Psychology, McGill University, Montreal, QC H3A 0G4, Canada; [c]Department of Psychology, University of Basel, 4055 Basel, Switzerland; and [d]Center for Neural Science, New York University, New York, NY 10003

Humans seamlessly infer the expanse of personality traits from others' facial appearance. These facial impressions are highly intercorrelated within a structure known as "face trait space." Research has extensively documented the facial features that underlie face impressions, thus outlining a bottom-up fixed architecture of face impressions, which cannot account for important ways impressions vary across perceivers. Classic theory in impression formation emphasized that perceivers use their lay conceptual beliefs about how personality traits correlate to form initial trait impressions, for instance, where trustworthiness of a target may inform impressions of their intelligence to the extent one believes the two traits are related. This considered, we explore the possibility that this lay "conceptual trait space"—how perceivers believe personality traits correlate in others—plays a role in face impressions, tethering face impressions to one another, thus shaping face trait space. In study 1, we found that conceptual and face trait space explain considerable variance in each other. In study 2, we found that participants with stronger conceptual associations between two traits judged those traits more similarly in faces. Importantly, using a face image classification task, we found in study 3 that participants with stronger conceptual associations between two traits used more similar facial features to make those two face trait impressions. Together, these findings suggest lay beliefs of how personality traits correlate may underlie trait impressions, and thus face trait space. This implies face impressions are not only derived bottom up from facial features, but also shaped by our conceptual beliefs.

face perception | impression formation | implicit personality theory | conceptual knowledge | social cognition

Humans naturally infer a broad range of personality traits from a face (1). From trustworthiness to creativity, we develop reliable impressions of others within seconds of seeing their face (2, 3). These face impressions influence our social behavior in situations as meaningful as election outcomes (4) and criminal sentencing (5).

Extensive research has documented how individual trait impressions are derived from morphological features of a face, for instance, that we infer both trustworthiness and submissiveness from babyfacedness (6). Naturally following, a central feature of face impressions is their highly intercorrelated structure (i.e., "face trait space"), in which each trait impression is correlated with many others (1). Thus, current perspectives explain face impressions as derived by specific facial features, and face trait space as emergent from the degree to which different trait impressions share a similar featural basis (e.g., kindness and submissiveness also relate to babyfacedness, and thus both correlate with trustworthiness; ref. 1). While such approaches have been highly valuable, they have tended to focus on a fixed architecture underlying face trait space, comprised of either two (1) or three (7) core dimensions that are commonly assumed to not change across perceivers.

In this article, we propose that face impressions, and thus their correlations (face trait space), are further structured by perceiver lay theories of others' personalities. Specifically, we propose that face impressions (e.g., intelligence) are also derived from the perception of other traits in a face (e.g., trustworthiness), insofar as a perceiver believes those two traits tend to correlate in other people. For example, a perceiver who believes the concept of trustworthiness is more related to the concept of intelligence

may see a trustworthy face as more intelligent. Research has long demonstrated that people hold rich lay conceptual associations of how they believe personality traits correlate in the world (in this article referred to as "conceptual trait space"; refs. 8–10). A common conceptual trait space has echoed throughout data-driven social perception research, where it has been long noted that a similar structure emerges across impression domains (face impressions, familiar person knowledge, stereotype content; refs. 1 and 11–15). Classic theory in person perception emphasized the role of this conceptual trait space in shaping initial impressions (i.e., lay, or implicit personality theories; ref. 16). For example, in seminal research of these questions, Asch (17) noted of his findings, "If a man is intelligent, this has an effect on the way in which we perceive his playfulness, happiness, friendliness" (p. 264). However, to our knowledge, such insights have not been directly applied or tested in understanding trait impressions of faces (although overlap in conceptual and face trait spaces has been observed toward romantic partner preferences; ref. 18). If a perceiver's conceptual associations in part help scaffold face trait space, this may further formal models of face impressions generally, and an important implication would be that face trait space is dynamic across perceivers rather than representing any single fixed architecture (9).

Across several studies, we describe evidence that perceivers' beliefs in trait associations, or conceptual trait space, relate to their impressions of faces and in turn the structure of faces' trait space. First, we demonstrate broadly that face trait space reflects conceptual trait space, finding substantial overlap between the two (study 1). Second, we find that perceivers' unique conceptual

Significance

Current theory of face-based trait impressions focuses on their foundation in facial morphology, from which emerges a correlation structure of face impressions due to shared feature dependence, "face trait space." Here, we proposed that perceivers' lay conceptual beliefs about how personality traits correlate structure their face impressions. We demonstrate that "conceptual trait space" explains a substantial portion of variance in face trait space. Further, we find that perceivers who believe any set of personality traits (e.g., trustworthiness, intelligence) is more correlated in others use more similar facial features when making impressions of those traits. These findings suggest lay conceptual beliefs about personality play a crucial role in face-based trait impressions and may underlie both their similarities and differences across perceivers.

trait associations are related to the correlations of their individual face impression judgments (studies 2 and 3). Lastly, we find perceivers' conceptual associations are related to the featural face space that underlies their impressions, which manifests in how they subjectively perceive individual traits in the first place (study 3). For all studies, all data and code are publicly available via the Open Science Framework (https://osf.io/z23kf/).

Results

Study 1. Given relatively common conceptual (19) and face (3) trait spaces among perceivers, they should show substantial overlap with one another on average if perceiver lay theories of personality shape their face impressions. It is possible that face trait space and conceptual trait space would not match. For instance, one can imagine the belief that dominant people are intelligent, responsible, and outgoing, yet the facial cues that give rise to dominance impressions may not give rise to intelligence impressions (1). These spaces could organize themselves by any number of factors that could structure trait concepts (e.g., valence, such as in a halo effect; ref. 11). Therefore, it is important to directly assess the correspondence of conceptual and face trait spaces. We first sought to empirically measure conceptual trait space (of the 13 traits used to estimate seminal models of face trait space; ref. 1), and assess whether face trait space reflects its structure. To do so, we used representational similarity analysis (RSA) (20), a powerful technique to assess similarity in such multivariate spaces (8, 9). In this technique, each trait space is represented as a similarity (i.e., correlation) matrix (pairwise relations of all traits to each other; Fig. 1 *A* and *B*), and then flattened into a vector of the unique pairwise similarity values between each trait. Because traits within a single matrix were measured on the same scale, similarities within each matrix were calculated using the standard distance metric of Pearson correlation, specifically, the pairwise correlations of trait judgments made of faces (*Materials and Methods*). But because raw values in different matrices have different meanings, we assessed the

correspondence between separate matrices (e.g., conceptual and face trait space) using Spearman rank correlation, which uses rank order rather than raw values (i.e., Pearson correlations) to estimate relationships between distances in the two spaces (20).

We measured conceptual and face trait space in two separate samples of participants. Each trait space was measured within a set of 13 personality traits used in seminal work quantifying face trait space: "aggressive," "caring," "confident," "dominant," "egotistic," "emotionally stable," "intelligent," "mean," "responsible," "sociable," "trustworthy," "unhappy," and "weird" (1). Similarities in the conceptual trait matrix were calculated using a straightforward pairwise similarity rating: the average degree to which participants believed each unique pairwise combination of personality traits are interrelated in other people ($n = 113$; e.g., trustworthy–dominant pair: "If someone is trustworthy, how likely are they to be dominant?") (Fig. 1B, *Top*) (*Materials and Methods*). To estimate face trait space, 90 different faces were rated by participants on each of the 13 trait stimuli ($n = 415$; each participant randomly assigned to one of the 13 trait stimuli) (Fig. 1B, *Bottom*). Indeed, as hypothesized, the conceptual and face similarity matrices explained a substantial amount of variance in one another {Spearman $\rho(76) = 0.82$, $\rho^2 = 0.67$, $P < 0.0001$, 95% CI = [0.74, 0.88]; Fig. 1C}. These results suggest that, when any two traits (e.g., caring and intelligent) are deemed more correlated in others, judgments of those traits in other people's faces exhibit a corresponding similarity or dissimilarity.

We replicated this relationship between conceptual and face trait matrices with a different face trait space, using trait judgment data from the original research defining the face trait space (1). (Note that egotistic was removed in this analysis, as it was not present in this specific dataset.) Indeed, a near-identical significant correlation between the conceptual and face trait model replicated this finding {Spearman $\rho(64) = 0.84$, $\rho^2 = 0.71$, $P < 0.0001$, 95% CI = [0.75, 0.90]}. Together, these results provide evidence for a strong correspondence between conceptual trait space and face trait space,

Fig. 1. Comparison of conceptual and face trait spaces. In study 1, we quantitatively assess the correspondence in structure of conceptual and face trait space. (*A*) Illustration of conceptual (*Top*) and face trait space models (*Bottom*) with multidimensional scaling (MDS). In our analysis, we tested correspondence of each trait space by the Spearman correlation of unique values above the diagonal of their similarity matrices (*B*; conceptual, i.e., "How likely is a person with one trait likely to have the other"; and face, i.e., how correlated are face impressions of one trait with another). Emo-stable, emotionally stable. (*C*) Analyses indicated the trait spaces overlap in structure substantially (13), Spearman $\rho(76) = 0.82$, $P < 0.0001$. Although the analysis was carried out using Spearman correlation, for illustrative purposes only Pearson correlation is depicted. MDS plots are organized by *k*-means clustering within each trait space, whereas both similarity matrices are sorted by the *k*-means clustering solution of the conceptual matrix for comparability.

consistent with a long history of research suggesting this correspondence (13–15, 18).

Study 2. Study 1 provides evidence that face trait space shares considerable structure with conceptual trait space (13). However, if conceptual associations play a role in shaping face trait space, perceivers' own face trait space should reflect their personal beliefs in how traits are conceptually associated. Thus, while conceptual and face trait spaces were estimated on average across subjects in study 1, study 2 accounted for between-subject differences in trait associations, assessing the relationship between perceivers' idiosyncratic conceptual and face trait spaces ($n = 206$). This question is an important step in addressing whether perceivers' own conceptual trait associations influence their face impressions. By current perspectives (1, 6), overgeneralized facial cues (e.g., resting smile resemblance of a face) activate specific trait concepts (e.g., trustworthiness) identically across perceivers, due to adaptive associations between traits and those overgeneralized cues, and regardless of perceivers' conceptual association between the cue-related trait impression (e.g., trustworthiness) and other trait impressions made from the same face (e.g., dominance, creativity). Such perspectives do not predict that face impressions would relate to individual differences in conceptual associations, whereas our account does indeed predict this.

Each participant was randomly assigned to one unique pair from a subset of the pairwise combinations in study 1: assertive, caring, competent, creative, self-disciplined, and trustworthy. (Due to practical limitations in measurement, note this looks through a pinhole at this process, only investigating single trait pairs per subject, rather than measuring the entirety of their trait spaces needed to acquire a full picture of this process.) Participants evaluated faces on both assigned traits, then later provided a conceptual similarity judgment between those traits, as in study 1. Therefore, in this study, participants served as the unit of analysis, with a score for their conceptual and face trait similarity. To test our hypothesis, we correlated participants' idiosyncratic face and conceptual trait similarities. Participants' conceptual similarity rating for a given trait pair was correlated with how similar those traits were judged in faces {Spearman $\rho(204) = 0.34$, $\rho^2 = 0.12$, $P < 0.0001$, 95% CI = [0.21, 0.46]} (Fig.

2A). These findings demonstrate a correspondence between how similar a participant idiosyncratically deems two traits and how similarly the participant judges those traits in others' faces. Thus, the results replicate and extend those of study 1, documenting correspondence between conceptual and face trait spaces on an individual level.

Study 3. We have seen that conceptual trait space and face trait space explain considerable variance in one another (study 1) and further explain individual differences in each other (study 2). These findings have testable implications for face impressions. If two different trait impressions are more or less correlated with one another, the facial features that typically evoke those impressions are likely to shift toward or away from one another, fundamentally altering the featural space underlying face impressions. In other words, perceivers who differ in the degree of conceptual association between traits would "see" these traits differently in faces. For instance, someone who believes agreeable people are often open to experience may make both impressions from faces based on more similar visual features. Someone who does not think agreeable people are often open to experience, on the other hand, may make both impressions based on less similar features.

To test this possibility, we applied a recently advanced reverse-correlation technique, which allowed us to estimate the facial features underlying participants' perceptions of traits in a data-driven manner (21). Using this technique, we obtained a featural vector in face space that represents each participant's visual representation of each trait. Thus, we estimated the perceived visual similarity of different traits in faces for each participant. Identical to study 2, we then tested whether a participant's idiosyncratic conceptual similarity between any two traits related to the visual similarity in features that evoke those specific traits for the participant. Each participant ($n = 185$) was randomly assigned to one unique pair from the unique pairwise combinations of the Big Five factor personality traits: "agreeable," "conscientious," "extroverted," "neurotic," and "open to experience." These traits were used to increase generalization of the findings of studies 1 and 2 and also given prior success in deriving these traits within the statistical face model we used (21). Participants performed a forced choice image classification task (e.g., ref. 22) for each trait assigned, then later
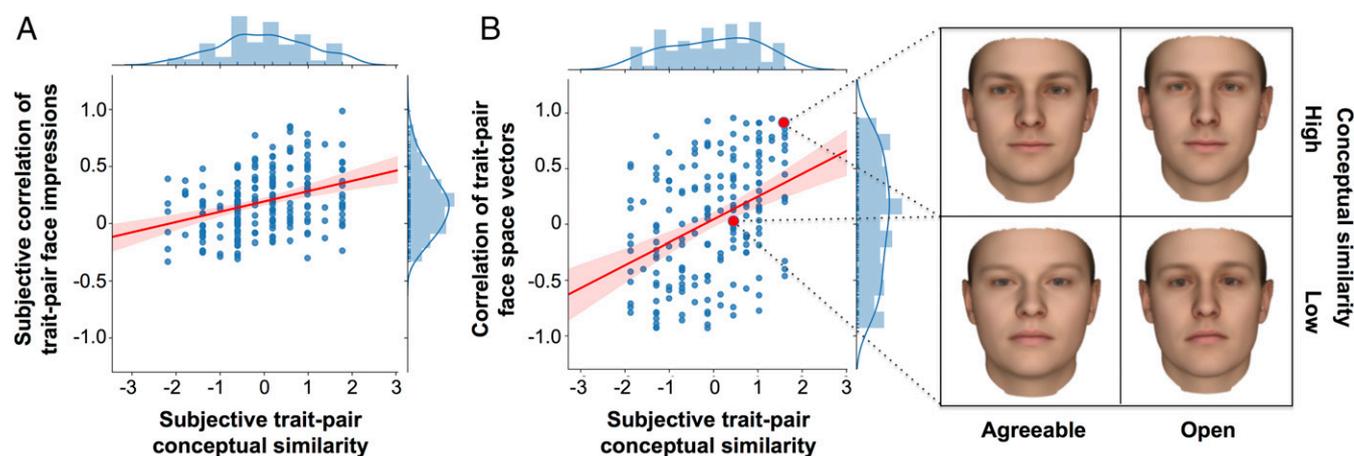

Fig. 2. Conceptual trait associations relate to visual similarity in facial features used for trait impressions. If lay conceptual beliefs about how personality traits correlate shape face impressions, perceivers' who believe two traits are more related (e.g., "agreeableness" related to "openness") should infer a trait from a face (e.g., agreeableness) to the extent they infer the other trait simultaneously from that face (e.g., openness), and thus see those traits more similarly in faces (e.g., illustration in B, Right). In study 2, we found participants who believed two personality traits were more correlated in others (e.g., "agreeable people are often open") also judged faces along those two traits more similarly (e.g., judged faces they perceived agreeable to also be open), Spearman $\rho(204) = 0.34$, $P < 0.0001$ (A). In study 3, participants with stronger conceptual associations between two traits (e.g., "agreeable people are often open") also used similar facial features to make those trait impressions of faces (e.g., facial features underlying agreeableness impressions were more similar to those underlying openness impressions; measured via image classification task), Spearman $\rho(183) = 0.40$, $P < 0.0001$ (B, Left). Although the analysis was carried out using Spearman correlation, for illustrative purposes only Pearson correlation is depicted. In B, Right, we also present two example participants to illustrate these findings, where a participant with high conceptual associations between agreeableness and openness (Top) sees those traits in faces more similarly than a participant low in that association (Bottom).

provided their idiosyncratic conceptual similarity rating between those traits. Accordingly, our data included each participant as the unit of analysis as in study 2, with a score for their conceptual and face trait similarity. Consistent with our hypotheses, a participant's conceptual similarity between two traits was correlated with the visual similarity in facial features associated with those traits {Spearman $\rho(183) = 0.40$, $\rho^2 = 0.16$, $P < 0.0001$, 95% CI = [0.27, 0.51]} (Fig. 2B). These findings show that the extent to which the visual features underlying each trait impression are more or less similar to those of other trait impressions relates to perceivers' own conceptual association between those traits. We illustrate this in Fig. 2B, in which we present the agreeable and open-to-experience classification images produced from two individual participant responses. For example, a participant who deems agreeableness and openness to be more conceptually related tends to see these traits as visually more similar in people's faces (i.e., uses similar features to make these impressions; see Fig. 2B).

Discussion

Together, our findings suggest that perceiver lay theories of personality may play an important role in face-based trait impressions. First, we found that conceptual trait space and face trait space explain a considerable amount of variance in each other (study 1). The relationship between conceptual trait associations and face trait associations is further evidenced by our findings that face impression judgments correlate within perceivers to the degree they believe those traits are more similar conceptually (study 2). Lastly, we found that conceptual trait associations predict the visual features perceivers use to infer those traits in others' faces. Thus, our findings provide correlational evidence suggesting that face impressions (e.g., intelligence) are partly derived from one another (e.g., trustworthiness), to the extent perceivers believe those traits are correlated in other people.

The current results provide several important contributions to theories of face impressions. The role of conceptual trait associations in face impression processes adds a crucial top-down layer to what have been predominately feature-driven bottom-up models (1, 6). If face impressions are derived from one another by way of their conceptual associations, this process may explain considerable correspondence in the structure of face impressions across perceivers (study 1; refs. 1 and 3), given similar correspondence in conceptual trait associations across perceivers (19). Above and beyond this commonality, this process may explain important individual differences in perceivers' face impressions and trait space (studies 2 and 3), to the extent their conceptual trait associations vary. As such, the findings bolster recent proposals arguing that face trait space may reflect a dynamic integration of not only intrinsic facial-feature covariation but also conceptual associations, stereotypes, and other social cognitive factors (9). Interestingly, the notion that individual differences in conceptual associations between traits shapes perceptions comports well with seminal person perception research that posited a role of "implicit personality theory" in nonface trait impressions (16, 17). The results therefore suggest that these classic insights with respect to general impression-formation patterns (outside of face perception) may apply to face-based trait impressions as well.

A common correlated structure of trait impressions has been observed not only in face impressions, but also in person knowledge and group-level stereotypes (1, 11, 12). This structure extends further to explain mental state inferences (23), as well as neural representations during social perception (24). That perceptions across domains share such similar structure is striking, and perhaps telling of a common cognitive basis for correlated social perceptions (13–15). Future research could directly investigate the role of conceptual trait spaces in shaping the structure of person perception in other domains, such as abstract representations of others (e.g., outside the domain of face evaluation; ref. 11) and social groups (12), including the possibility of empirically connecting these various spaces together. Indeed, RSA approaches (20) have been leveraged to assess

correspondences across various domains, such as actual personality (25) and social categorization (26). In addition, understanding the contribution of perceiver conceptual trait associations to social perception across these domains may be paramount to understanding real-world social behavior that is quite consequential. Dimensions of both face impressions and group stereotypes are highly consequential, in situations as serious as election outcomes (4, 27) and criminal sentencing (5, 28). Future research should assess whether important individual and cultural differences in conceptual trait space alter critical social decisions.

With respect to these dimensions, the results may provide a parsimonious explanation for cases in which their correlations may cease to be independent and shift. In one example, trait impressions of less familiar others may be more intercorrelated and lower dimensional than those of familiar others (29, 30). It may be the case that perceivers rely more on their conceptual trait space, in which trait judgments are highly correlated, to make impressions of unfamiliar others when more specific person knowledge is unavailable. For example, additional information about targets allows trait dimensions of sociability and morality, typically linked to one another (1, 12), to become orthogonal (31). This account could also generalize to explain models of trait impressions in intergroup contexts. For instance, use of a conceptual trait space to make wide personality inferences toward unfamiliar outgroup members may underlie systematically biased (32–34) and therefore homogenous trait impressions (35). However, increased information about targets may disengage use of the conceptual trait space (i.e., individuation; ref. 33). Another notable example is the more negative relationship between trustworthiness and dominance impressions of female compared with male faces (36), presumably due to stereotypes linking female likability with submissiveness (37). Our findings suggest that unique conceptual trait spaces, such as when considering different social groups (e.g., conceptual associations between traits when regarding females vs. males), may lead to differential associations between face impressions. Future research could measure shifts in conceptual trait space in different social contexts to assess whether variations in face trait and group-level trait space emerge from a conceptual basis.

There are important limitations of the current work. Most notably, the correlational nature of our design precludes any strong inference about the causal impact of conceptual knowledge on face trait space. Alternative possibilities exist, including face impressions shaping conceptual trait space. At face value, it seems unlikely that individual differences in face impression correlations (due to mere featural processing of the same face stimuli) could exert such a consistent influence on participant conceptual associations between personality traits. This is especially the case, given perceivers would have to track whether impressions of faces from one task somehow reflected those in the second separate task; and there is a considerable lack of awareness concerning which features underlie perceivers' judgments (2, 38–40). However, our current data cannot exclude these possibilities. Future research should seek causal evidence of the influence of conceptual knowledge on face trait space by manipulating conceptual knowledge directly.

Another noteworthy limitation is the use of language—trait concept terms such as trustworthiness—to measure both face impressions and conceptual associations. This issue has been central to longstanding debates concerning the origins of lay personality theory models, in which researchers have debated whether measured trait concept associations are merely semantic in nature, rather than underlain by beliefs about actual personality traits of others (for a review, see ref. 16). If perceivers' trait-term semantic associations (e.g., believing the words "kind" and "sociable" mean the same thing) are all that is behind their conceptual and face trait associations, similarity in conceptual and face trait spaces may be an artifact of language and uninteresting for understanding social behavior. Speaking against this possibility, many researchers have found evidence that trait concept correlations are independent of semantic features and argued semantic explanations do not obviate socially meaningful and consequential trait relations (41, 42). Nonetheless, such

ruling out has not been applied in the current domain of face impressions, which future research should evaluate in this context. Future research could examine whether the significance of a trait impression changes, for instance whether conceptual shifts in intelligence impressions impact its affective (e.g., evaluative priming) or behavioral (e.g., hiring decisions) consequences for perceivers.

In conclusion, we found that lay conceptions of personality traits are strongly related to trait impressions based on other people's facial appearance. The common structure that emerges across perceivers in face impressions (1, 3) has considerable resemblance to commonly shared conceptual trait structure (11). Beyond any such shared structure, individual differences in perceivers' conceptual trait associations are related to the unique structure of their face impressions and the features that underlie them. Together, these findings suggest that the way we infer personality traits from faces is not only determined by the physical appearance of a face, but also by our own lay conceptual beliefs regarding the personality of others.

Materials and Methods

Data, analysis code, and results are all available and hosted by the Open Science Framework (https://osf.io/z23kf/). Data may be downloaded, and results reproduced via Jupyter notebooks available in the repository.

Study 1.
Participants.

Face trait space. We collected face impression data from 415 subjects via Amazon Mechanical Turk (demographic data missing for 1 subject; all US residents; all primary English speakers; M_{age} = 34.23 y, SD_{age} = 12.27 y; 260 female, 146 male, 2 other, 5 decline; 316 White, 33 Black, 28 Asian, 38 other). Participants were randomly assigned to evaluate one personality trait in all face stimuli and therefore divided roughly equally between all 13 personality trait conditions (~32 participants per trait condition). Subjects were financially compensated for their participation, and they gave informed consent. This experiment was approved by the University Committee on Activities Involving Human Subjects at New York University.

Conceptual trait space. We collected conceptual trait association data from 113 subjects via Amazon Mechanical Turk (demographic data missing for 1 subject; all US residents; all primary English speakers; M_{age} = 36.34 y, SD_{age} = 11.14 y; 72 female, 40 male; all White). Subjects were financially compensated for their participation, and they gave informed consent. This experiment was approved by the University Committee on Activities Involving Human Subjects at New York University.
Stimuli.

Face stimuli. All stimuli were taken from the Chicago Face Database (43). Face stimuli included 90 portrait photographs of young white male individuals with neutral facial expressions. These stimuli were also used in study 2. A secondary analysis looked at a face trait similarity model derived from seminal work in face trait space measurement. In this study (1), 66 faces (female and male) from the Karolinska Directed Emotional Faces face database (44) were rated on each trait (other than "egotistic"; 1–9 Likert-type scale; e.g., 1 = "not at all trustworthy,"; 9 = "extremely trustworthy"). See the original publication (1) for additional details (data available upon request from the authors' web database; https://tlab.princeton.edu/databases/).

Personality trait stimuli. We chose 13 personality traits that independent groups of participants evaluated in faces and in conceptual similarity. These traits were those used in the seminal work assessing face trait space (1). In this work, these traits were chosen as those unique but also spontaneously elicited during face impressions (with the exception of dominance, which was included by the researchers). These traits included: aggressive, caring, confident, dominant, egotistic, emotionally stable, intelligent, mean, responsible, sociable, trustworthy, unhappy, and weird.
Protocol. See *SI Appendix* for detailed task instructions.

Face trait space task. Participants were informed they would partake in a study examining how people perceive others. Each participant was randomly assigned to evaluate only 1 of the 13 personality trait stimuli in faces. In the task, participants rated each of the 90 face stimuli on the personality trait they were assigned (1–7 Likert-type scale; e.g., 1 = "very untrustworthy," 4 = "neutral," 7 = "very trustworthy"). Following the face trait rating task, participants completed a general demographics survey and completed the experiment.

Conceptual trait space task. Participants were informed they would partake in a study on how different personality traits correlate in the world. Participants evaluated the conceptual relationship of each trait pair in the 13 trait stimuli (1–7 Likert-type scale, 1 = "not at all likely," 4 = "neutral," 7 = "very likely"), presented in both orders (e.g., "trustworthy – dominant" and "dominant – trustworthy"). Therefore, there were a total of 156 trials for each participant [P(13,2) = 156]. Following the face trait rating task, participants completed a general demographics survey and completed the experiment.
Data preparation and analysis. All analyses were conducted with scientific and statistical libraries in Python. No subjects were removed from these data before analysis. To assess whether face trait space reflects conceptual trait space, we applied a quantitative method from systems neuroscience, RSA (20). As a straightforward explanation, this analysis measured the correlation between trait-pair similarity matrices as measured in the face trait and conceptual trait tasks. An intuitive description of this process is to correlate the unique values of two different similarity matrices together, assessing the correspondence between the two similarity matrices. Therefore, we may assess whether the similarities of face trait judgments reflect the pattern of how similar those traits are conceptually conceived. See a detailed explanation of RSA in *SI Appendix*.

Study 2.
Participants. We collected face impression data from 206 subjects via Amazon Mechanical Turk (original *n* = 213; 2 subjects dropped out due to task incompletion; 5 subjects dropped out due to failure to follow task instructions; all US residents; all primary English speakers; M_{age} = 29.78 y, SD_{age} = 6.81 y; 102 female, 65 male, 1 decline; gender data from 38 participants missing due to a data collection error; 160 White, 17 Black, 9 Asian, 20 other). Subjects were financially compensated for their participation, and they gave informed consent. This experiment was approved by the University Committee on Activities Involving Human Subjects at New York University.
Stimuli.

Face stimuli. Face stimuli were identical to those collected in our data in study 1 (*Materials and Methods*, Study 1).

Personality trait stimuli. We chose a diverse set of trait stimuli somewhat deviating from those in study 1 to assess generalizability. Trait stimuli included: assertive, caring, competent, creative, self-disciplined, and trustworthy. We used all pairwise combinations of these trait pairs (for a total of 15 unique possible trait pairs). Participants were randomly assigned to 1 of the 15 total trait-pair combinations.
Protocol. Both face trait and conceptual trait tasks were largely identical in design within themselves to those in previous studies (*Materials and Methods*, Study 1). A major distinction is that in study 2, each participant both provided face trait and conceptual trait data. Each participant was randomly assigned to 1 of 15 trait pairs (the unique combinations of six trait stimuli: assertive, caring, competent, creative, self-disciplined, and trustworthy). First, participants evaluated all face stimuli on both assigned traits. They evaluated all stimuli on one trait first, followed by the other. The order of which trait was first evaluated was randomly determined per subject. In total, participants therefore completed 180 trials of face impressions. From these data, we were able to measure the correlation of face impressions within each subject. Second, participants provided conceptual trait association ratings for their assigned trait pair. As participants only evaluated the similarity of two traits to one another (compared with the many trait pairs in study 1), there were only two trials in the conceptual trait task. Instructions and item design were identical to those used in study 1. Following these tasks, participants completed a general demographics survey.
Data preparation and analysis. In study 2, we ask whether the amount to which each perceiver associates two trait concepts relates to the correlation between those trait impressions in faces. That is, we intended to test whether perceivers with weaker/stronger conceptual trait associations also show more weakly/strongly correlated face impressions. To do so, within each perceiver, we calculated two variables: their conceptual and face trait associations (*SI Appendix*). To test our hypothesis, we calculated the Spearman correlation between participant face trait and conceptual trait associations (Spearman correlation used so as to not assume a strictly linear relationship between distances in the two spaces) (20). Analyses were conducted across trait-pair terms, to assess the tendency of conceptual trait associations to relate to face impression correlations, across trait pairs in general.

Study 3.
Participants. We collected face trait image classification data from 186 subjects via Amazon Mechanical Turk (original *n* = 194; 9 subjects removed due to

task incompletion; all US residents; all primary English speakers; $M_{age} = 33.89$ y, age data for one subject missing, $SD_{age} = 8.6$ y; 113 female, 72 male, 1 decline; 139 White, 21 Black, 11 Asian, 15 other). Subjects were financially compensated for their participation, and they gave informed consent. This experiment was approved by the University Committee on Activities Involving Human Subjects at New York University.

Stimuli.

Face stimuli. First, we created an average face from 100 female and 100 male faces from the Basel Face Model (45). Within the shape and the color space spanned by these 200 faces, we created 100 vectors randomly varying face shape and 100 vectors randomly varying face color. Separately applying these 200 vectors to the average face in both positive and negative direction resulted in 200 pairs of faces or 200 classification trials, respectively.

Personality trait stimuli. Personality trait stimuli included the Big Five personality traits (agreeable, conscientious, extroverted, neurotic, and open to experience), due to their successful use in prior work with this statistical face manipulation technique (21). Furthermore, these new trait stimuli allowed us to even further diversify our trait stimuli to strengthen inferences of generalizability. We used all pairwise combinations of these trait pairs (for a total of 10 unique possible trait pairs). Participants were randomly assigned to 1 of the 10 total trait-pair combinations.

Protocol. The overall structure of study 3 was similar to the structure used in study 2. Each participant provided both face trait and conceptual trait data. Participants were randomly assigned to 1 of the 10 trait-pair permutations (i.e., one of the pairwise combination of the Big Five traits, varying in order by which trait was listed first to counterbalance the task below). Each participant completed four image classification tasks. They first performed a shape and a color task for the first trait they were assigned to, followed by a shape and a color task for the second trait they were assigned to. All four tasks comprised 100 trials. In each trial, participants were presented with two faces horizontally adjacent to one another on the same page (i.e., random vector applied to the average face in positive direction and in negative direction), and asked to indicate which of the two faces looks more extreme regarding the trait in question (e.g., which face looks more agreeable). Following the image classification task, participants provided conceptual trait association ratings for trait pairs assigned. This task was identical to that in study 2. Lastly, participants completed a general demographics survey.

Data preparation and analysis. In study 3, we asked whether the amount to which each perceiver associates two trait concepts is related to the correlation between those traits' face space feature vectors (i.e., "face trait vectors") estimated from the image classification task. That is, we tested whether perceivers with weaker/stronger conceptual trait associations actually see traits less/more similarly in faces. Within each perceiver, we calculated two variables: their face trait vectors' correlation and conceptual trait associations (*SI Appendix*). To test our hypothesis, we calculated the Spearman correlation between participant face trait vectors and conceptual trait associations (Spearman correlation used so as to not assume a strictly linear relationship between distances in the two spaces; ref. 20). Analyses were conducted across trait-pair terms, to assess the tendency of conceptual trait associations to predict face trait vector correlations, across trait pairs in general.

1. Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. *Proc Natl Acad Sci USA* 105:11087–11092.
2. Todorov A, Pakrashi M, Oosterhof NN (2009) Evaluating faces on trustworthiness after minimal time exposure. *Soc Cogn* 27:813–833.
3. Hehman E, Sutherland CA, Flake JK, Slepian ML (2017) The unique contributions of perceiver and target characteristics in person perception. *J Pers Soc Psychol* 113: 513–529.
4. Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence from faces predict election outcomes. *Science* 308:1623–1626.
5. Wilson JP, Rule NO (2015) Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychol Sci* 26:1325–1331.
6. Zebrowitz LA, Montepare JM (2008) Social psychological face perception: Why appearance matters. *Soc Personal Psychol Compass* 2:1497.
7. Vernon RJ, Sutherland CA, Young AW, Hartley T (2014) Modeling first impressions from highly variable facial images. *Proc Natl Acad Sci USA* 111:E3353–E3361.
8. Tamir DI, Thornton MA (2018) Modeling the predictive social mind. *Trends Cogn Sci* 22:201–212.
9. Stolier RM, Hehman E, Freeman JB (2018) A dynamic structure of social trait space. *Trends Cogn Sci* 22:197–200.
10. Osgood CE (1952) The nature and measurement of meaning. *Psychol Bull* 49:197–237.
11. Rosenberg S, Nelson C, Vivekananthan PS (1968) A multidimensional approach to the structure of personality impressions. *J Pers Soc Psychol* 9:283–294.
12. Fiske ST, Cuddy AJ, Glick P, Xu J (2002) A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J Pers Soc Psychol* 82:878–902.
13. Todorov A, Said CP, Engel AD, Oosterhof NN (2008) Understanding evaluation of faces on social dimensions. *Trends Cogn Sci* 12:455–460.
14. Fiske ST, Cuddy AJ, Glick P (2007) Universal dimensions of social cognition: Warmth and competence. *Trends Cogn Sci* 11:77–83.
15. Oldmeadow JA, Sutherland CA, Young AW (2013) Facial stereotype visualization through image averaging. *Soc Psychol Personal Sci* 4:615–623.
16. Schneider DJ (1973) Implicit personality theory: A vreview. *Psychol Bull* 79:294–309.
17. Asch SE (1946) Forming impressions of personality. *J Abnorm Psychol* 41:258–290.
18. South Palomares JK, Sutherland CA, Young AW (2017) Facial first impressions and partner preference models: Comparable or distinct underlying structures? *Br J Psychol* 109:538–563.
19. Kuusinen J (1969) Factorial invariance of personality ratings. *Scand J Psychol* 10:33–44.
20. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis:– Connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
21. Walker M, Vetter T (2016) Changing the personality of a face: Perceived big two and big five personality factors modeled in real photographs. *J Pers Soc Psychol* 110: 609–624.
22. Dotsch R, Wigboldus DH, Langner O, van Knippenberg A (2008) Ethnic out-group faces are biased in the prejudiced mind. *Psychol Sci* 19:978–980.
23. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. *Science* 315:619.
24. Tamir DI, Thornton MA, Contreras JM, Mitchell JP (2016) Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proc Natl Acad Sci USA* 113:194–199.
25. Lay CH, Jackson DN (1969) Analysis of the generality of trait-inferential relationships. *J Pers Soc Psychol* 12:12–21.
26. Stolier RM, Freeman JB (2016) Neural pattern similarity reveals the inherent intersection of social categories. *Nat Neurosci* 19:795–797.
27. Hehman E, Carpinella CM, Johnson KL, Leitner JB, Freeman JB (2014) Early processing of gendered facial cues predicts the electoral success of female politicians. *Soc Psychol Personal Sci* 5:815–824.
28. Eberhardt JL, Davies PG, Purdie-Vaughns VJ, Johnson SL (2006) Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychol Sci* 17:383–386.
29. Thornton MA, Mitchell JP (2017) Theories of person perception predict patterns of neural activity during mentalizing. *Cereb Cortex,* 1–16.
30. Koltuv BB (1962) Some characteristics of intrajudge trait intercorrelations. *Psychol Monogr* 76:1–33.
31. Brambilla M, Rusconi P, Sacchi S, Cherubini P (2011) Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *Eur J Soc Psychol* 41:135–143.
32. Kunda Z, Thagard P (1996) Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychol Rev* 103:284–308.
33. Fiske ST, Neuberg SL (1990) A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Adv Exp Soc Psychol* 23:1–74.
34. Brewer MB (1988) A dual process model of impression formation. *A Dual-Process Model of Impression Formation: Advances in Social Cognition*, eds Srull TK, Wyer RS (Erlbaum, Hillsdale, NJ), Vol 1, pp 1–36.
35. Quattrone GA, Jones EE (1980) The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *J Pers Soc Psychol* 38:141–152.
36. Sutherland CA, Young AW, Mootz CA, Oldmeadow JA (2015) Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *Br J Psychol* 106:186–208.
37. Glick P, Fiske ST (1996) The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *J Pers Soc Psychol* 70:491–512.
38. Freeman JB, Stolier RM, Ingbretsen ZA, Hehman EA (2014) Amygdala responsivity to high-level social information from unseen faces. *J Neurosci* 34:10573–10581.
39. Ambady N, Bernieri FJ, Richeson JA (2000) Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology* (Academic, San Diego), Vol 32, pp 201–271.
40. Tskhay KO, Rule NO (2013) Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Pers Soc Psychol Rev* 17:72–86.
41. Borkenau P (1992) Implicit personality theory and the five-factor model. *J Pers* 60:295–327.
42. Block J, Weiss DS, Thorne A (1979) How relevant is a semantic similarity interpretation of personality ratings? *J Pers Soc Psychol* 37:1055–1074.
43. Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: A free stimulus set of faces and norming data. *Behav Res Methods* 47:1122–1135.
44. Lundqvist D, Flykt A, Öhman A (1998) *The Karolinska Directed Emotional Faces—KDEF* (Karolinska Institute, Stockholm).
45. Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T (2009) A 3D face model for pose and illumination invariant face recognition. *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'09)*. Available at https://ieeexplore.ieee.org/document/5279762/. Accessed August 16, 2018.