This is a preprint of this manuscript, currently in press at *Nature Human Behaviour*, version dated November 27th, 2019 (date will be updated with preprint, prior versions documented by OSF). Project data and analysis materials are available online: https://osf.io/2uzsx/ .

Trait knowledge forms a common structure across social cognition

Ryan M. Stolier[1]*

Eric Hehman[2]

Jonathan B. Freeman[3,4]*

[1]Department of Psychology, Columbia University, New York, NY, USA

[2]Department of Psychology, McGill University, Montreal, QC, Canada

[3]Department of Psychology, New York University, New York, NY, USA

[4]Center for Neural Science, New York University, New York, NY, USA

Corresponding authors:
Ryan M. Stolier & Jonathan B. Freeman
Department of Psychology
Columbia University
406 Schermerhorn Hall
1190 Amsterdam Ave.
New York, NY 10027
Phone: (212) 851-9348
Email: rms2262@columbia.edu & jon.freeman@nyu.edu

Abstract

Researchers have noted the resemblance across core models of social cognition, in which trait inferences center on others' intentions and abilities (e.g., warmth, competence). Current views posit this common 'trait space' originates from the adaptive utility of the dimensions, predicting a relatively fixed and universal architecture. In contrast, we hypothesize perceivers learn conceptual knowledge of how traits correlate that shapes trait inferences similarly across domains (e.g., faces, person knowledge, stereotypes), from which a common trait space emerges. Here we show substantial overlap between the structures of perceivers' conceptual and social perceptual trait spaces, across perceptual domains (Studies 1-4), and that conceptual associations directly shape trait space (Study 5). Furthermore, we find evidence that conceptual trait space is learned from social perception and actual personality structure (Studies 6-7). Our findings suggest conceptual trait associations serve as a cornerstone in social perception, providing broad implications to the study of social behavior.

TRAIT KNOWLEDGE FORMS A COMMMON STRUCTURE ACROSS SOCIAL COGNITION

To navigate an exceptionally complex social world, we ascribe countless traits to one another. Yet, this sea of trait inferences cohere into a small set of dimensions comprising a 'trait space' in social cognition: most often two dimensions, concerning others' intentions (e.g., warmth, trustworthiness, communion) or capacity to enact those intentions (competence, dominance, agency; for review, see[1,2]). This trait space seems conspicuously similar across a variety of distinct domains in social cognition, such as first impressions from faces[3], knowledge of familiar others[4], and group stereotypes[5]. Thus, it has been theorized that social cognition has a fixed architecture structured around a set of universal dimensions, often interpreted to reflect that humans track intention and capability traits given their utility in guiding adaptive social behavior[2,6]. While such a process may explain why certain traits are central to trait dimensions (e.g., morality to the warmth dimension[7,8]), recent research has found substantial variation in the dimensionality of trait space[9], and it is still unclear why the countless traits (e.g., sociability, humor, neuroticism, liberalism) correlate along these dimensions as they do. Moreover, it is unclear whether the organization of trait inferences along low-level dimensions is merely an emergent property of social perception (e.g., tracking central traits of warmth and competence[2]), or plays a functional role in forming social perceptions and trait inferences in the first place.

Another possible explanation of a common trait space in social cognition is that perceivers may hold subjective conceptual knowledge of how personality traits correlate in others, which then guide trait inferences similarly across many social cognitive domains. For instance, perceivers may believe kind others are often intelligent, and thus judge a kind face, reputed other, or social group to also be intelligent. This would cause trait inferences to correlate similarly across social perception, thus producing a common trait space. Classic research regarding such conceptual associations (i.e., implicit personality theory[10]) has shown that they influence trait inferences and trait space during impression formation based upon vignettes[7,11].

Decades of personality research indicates personality traits are actually correlated along a lower set of dimensions (e.g., two-factor and five-factor solutions, such as the Big Five[11,12]). It may be the case that perceivers learn how personality correlates through various means, such as cultural transmission, direct observations, and interaction (for review, see[10,13]), and apply this knowledge to infer others' traits whether in pursuit of accuracy or due to the inevitable effects of associative processing. This inferential process may be analagous to how perceivers' cognitive models of mental state associations are applied to accurately predict a target's future mental states based upon their current one[14,15]. Thus, here we explore the possibility that perceivers conceptually learn actual trait associations, and use those learned conceptual associations when perceiving others.

In the present research, we extend this conceptually-driven stance on trait inferences to explain a commonality in trait space structure across social cognitive domains (including face impressions, familiar person knowledge, and group stereotypes; Studies 1 - 5; Fig. 1), which may be a by-product of the applying learned trait knowledge to form initial inferences (Studies 6 & 7). This process diverges from a universal account[2] in that the mechanism underlying trait space structure is not an evolved tendency for tracking functionally adaptive information, but rather reflects general conceptual knowledge about personality. Because such knowledge may differ across individuals depending on their experiences or learning, this perspective also provides a parsimonious account of emerging evidence that trait space changes across individual perceivers and social contexts[9,16-18]. We find evidence in support of this account across several studies. All stimuli, data, and analysis scripts (Python, R) are available on the OSF, from which results may be reproduced (https://osf.io/2uzsx/).

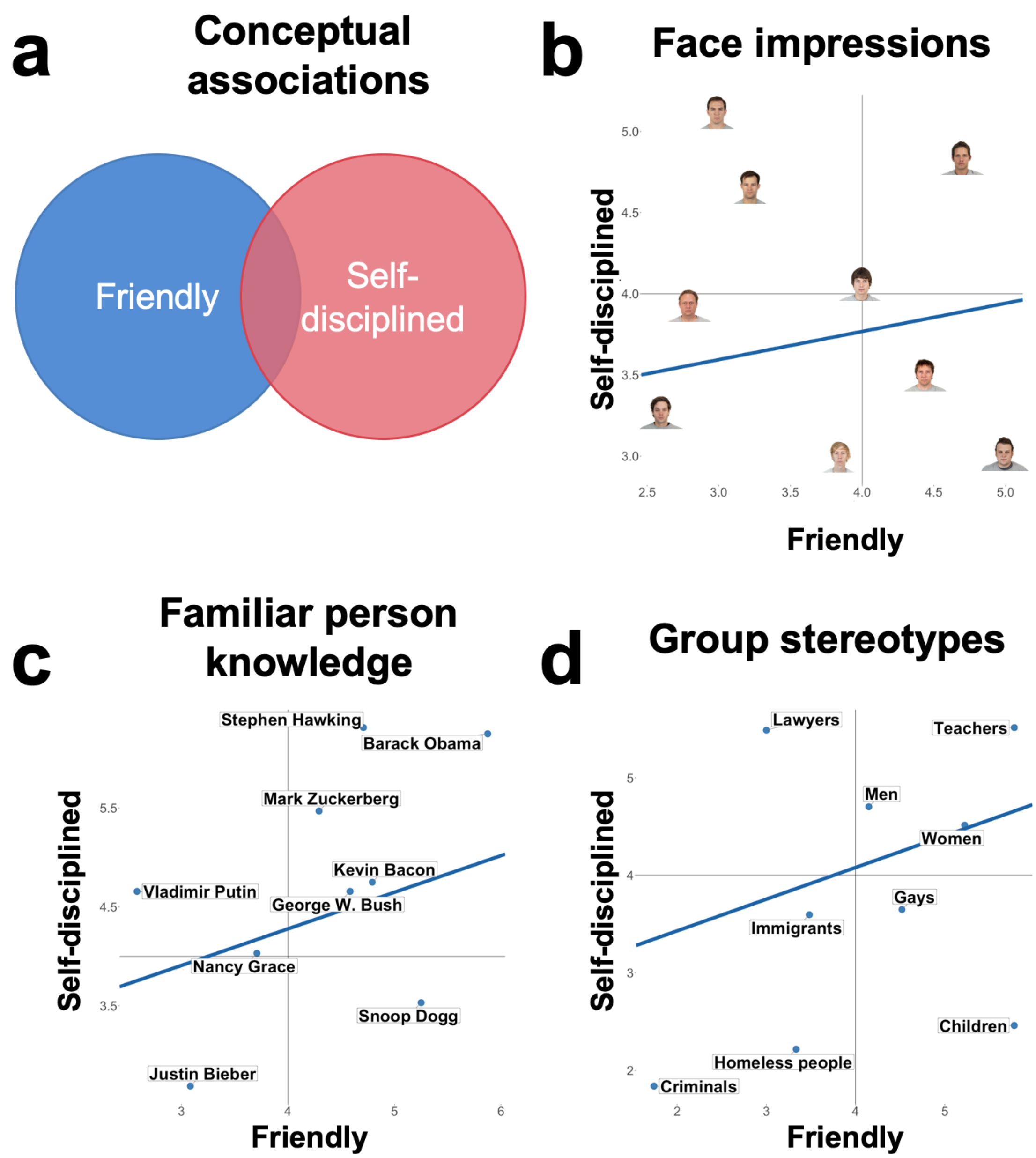


Figure 1. Conceptual illustration of theoretical and analytic approach. We theorize that conceptual trait associations (panel a) shape social perceptions across domains (panels b,c,d). Thus our analytic strategy across studies was to evaluate whether the pair-wise relationships between personality traits across these domains are similar. For instance, we observe that the conceptual associations between two personality traits, 'friendly' and 'self-disciplined' (panel a), is mirrored by the correlation of 'friendly' and 'self-disciplined' perceptions of faces (panel b), familiar others (panel c), and social groups (panel d). Please note this figure is for illustrative purposes. It is important to note analyses reported do not test the similarity in magnitude of trait-pair correlations between domains, but rather if the rank ordering of associated trait-pairs is similar between domains (see Methods). (Panel a is conceptually illustrative, whereas panels b, c, and d depict a subset of data from Study 1, 9 data points per panel. Several trait space and stimulus examples are provided as data points in panels b, c, and d.)

Results

Study 1

In Study 1, we compared models of various social perceptual trait spaces to a model of conceptual trait space (see Methods; Fig. 2). Distinct sets of participants reported their conceptual associations between traits (n = 116; e.g., 'Are kind people often intelligent?'), and impressions towards unfamiliar faces (for collection of social perceptual similarity matrices in Studies 1 and 2, sample size reported is total participant raters, where subsets this total reported sample rated each trait; n = 484; e.g., 'How kind/intelligent is this face?'), familiar famous and historical people (n = 503; e.g., 'How kind/intelligent is Barack Obama?'), and social groups (n = 488; e.g., 'How kind/intelligent are teachers?'). We looked at a trait space of 15 trait terms across domains, made up of three sub-traits per each of the big five factors of personality (see Methods; Fig. 2[19]). From these data, we computed a similarity matrix for each of these four domain models (Fig. 2a). Each matrix is a collection of all pair-wise 'similarities' in each domain, where similarity in the conceptual trait space matrix is the conceptual association between each trait-pair ('How likely are kind people to be intelligent?'; 1 – 7 Likert-type item), and in each perceptual trait space matrix is the pair-wise Pearson correlation between each trait inference (e.g., correlation of 'kind' and 'intelligent' face impressions). We then applied representational similarity analysis (RSA), testing the Spearman correlation between unique values in each matrix-pair (Spearman correlation used so as to not assume a strictly linear relationship between distances in the two spaces[20]). In effect, this tests whether the relative degree of correlation between trait-pairs across different trait spaces is the same. To account for structured dependency of matrices and allow inference towards our participants (given Studies 1, 2, and 7 collapse across subjects), additional analyses for all such RSA analyses are reported in the Supplementary Results.

All pairwise trait space matrix analyses are depicted in Figure 2 (panels b,c,d). We tested our hypothesis that a common social perceptual trait space may reflect a more domain-general conceptual trait space. Consistent with this hypothesis, we observed significant similarity between the conceptual trait space matrix and all social perceptual trait space matrices (Fig. 2b,c,d; conceptual matrix predicts: face trait space matrix, Spearman $\rho(103) = 0.796$, $\rho^2(103) = 0.634$, $p < 0.0001$; 95% CI = [0.713 , 0.857]; familiar person trait space matrix, Spearman $\rho(103) = 0.739$, $\rho^2(103) = 0.545$, $p < 0.0001$; 95% CI = [0.637 , 0.815]; social group trait space matrix, Spearman $\rho(103) = 0.779$, $\rho^2(103) = 0.606$, $p < 0.0001$; 95% CI = [0.690 , 0.844]). Additional analyses confirmed a strong correspondence among the three social perceptual trait matrices (see Supplementary Results). Moreover, RSA using a valence similarity matrix (Supplementary Figure 1) demonstrated that all reported effects occur above and beyond any effects due to valence alone. These findings provide evidence in support of our theoretical hypothesis that a domain-general conceptual trait space is reflected in a common social perceptual trait space seen across several domains.

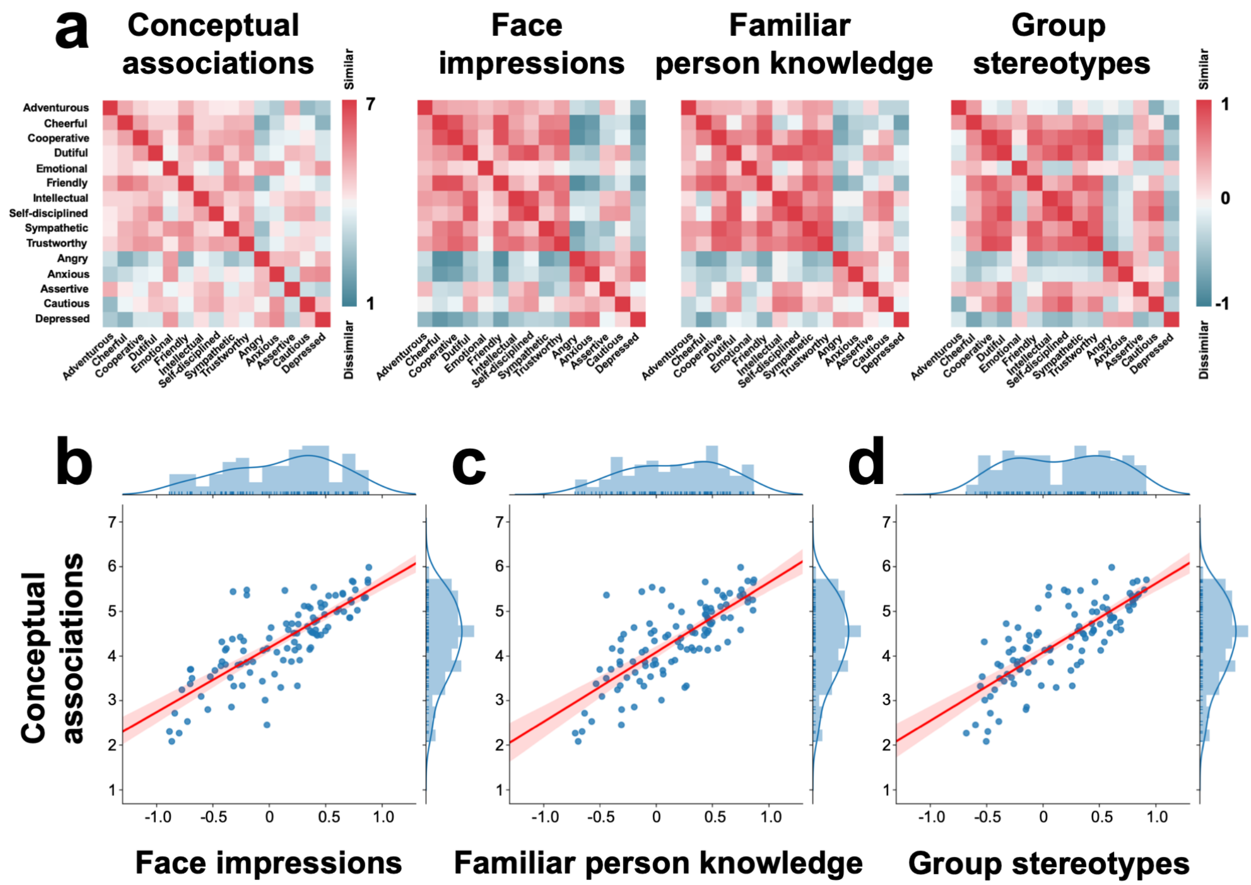


Figure 2. **Trait inferences across social cognition mirror conceptual knowledge.** First depicted are all trait space similarity matrices from Study 1 (panel a), each made of the pairwise similarity values between each trait-pair (plotted from dissimilar/blue to similar/red). Each matrix is sorted by the k-means cluster solution of the conceptual trait space matrix, as to most intuitively depict their similar structure. Each matrix was collected from a distinct task, set of stimuli, and set of participants. The results show that conceptual trait space (n = 116) is powerfully reflected in social perceptual trait spaces across domains (face impressions, panel b, n = 484, Spearman $\rho(103) = 0.796$, $\rho^2(103) = 0.634$, $p < 0.0001$; 95% CI = [0.713 , 0.857]; person knowledge, panel c, n = 503, Spearman $\rho(103) = 0.739$, $\rho^2(103) = 0.545$, $p < 0.0001$; 95% CI = [0.637 , 0.815]; group stereotypes, panel d, n = 488, Spearman $\rho(103) = 0.779$, $\rho^2(103) = 0.606$, $p < 0.0001$; 95% CI = [0.690 , 0.844]). Error ribbons display standard error of the estimate, and there are 105 trait-pairs as data points per panel. While Pearson correlations are plotted for ease of interpretation, statistical analyses were of rank ordered data points. In each plot, trait space matrices (panel a) are flattened into their unique pair-wise similarity values and plotted against one another (conceptual on the y-axis, panel a, left-most matrix; social perceptual matrices along the x-axes, panel a, right three matrices). Each data point is a trait-pair (e.g., ‘friendly’-‘self-disciplined’). In each comparison, as two traits become more associated in conceptual knowledge (y-axis), they become more correlated in trait inferences across domains (x-axis). This pattern is found in Study 1, in which trait terms were used in each task (e.g., ‘friendly’, ‘self-disciplined’), and in Study 2, in which different trait descriptors were used in each task (e.g., ‘friendly’ in the conceptual task, ‘likely’ to agree with others’ in the face task; see Study 2 results and Supplementary Figure 2).

Study 2

An important aspect of our theoretical perspective is that trait conceptual knowledge drives inferences regarding social cognition, which are used for understanding other people and predicting their behaviors. We may predict a ‘kind’ person who behaves affectionately to be ‘extroverted’ and socialize frequently. One alternative interpretation of the results is that the correlation of any two trait inferences, such as ‘kind’ and ‘extroverted’, is due merely to how participants find any two words synonymous in semantic meaning (for review, see[10,21]). To highlight the role of the trait concepts measured here as meaningful concepts that reflect perceivers’ differential predictions about human behavior[15], eliminate concerns regarding semantics, and provide a conservative conceptual replication of Study 1, in Study 2 we designed a set of tasks emphasizing traits as distinct concepts used to predict distinct behaviors in a substantive manner[15]. Rather than asking participants about the same trait terms across domains (e.g., ‘is this face kind?’ and ‘is this group kind?’), we used different items for each domain, which asked about the behavioral tendencies thought to underlie personality traits (e.g., instead of ‘kind’: ‘is this face likely to agree with others?’ and ‘is this social group likely to compliment others?’). We gathered several items to correspond uniquely to each trait. Thus we used behavioral tendency descriptions as proxies for traits for each of the different social perceptual domains, and compared the similarity matrices to the conceptual similarity matrix collected in Study 1 that used direct trait terms. Items were chosen from the NEOPI[19], given both its use of behavioral tendency descriptions to collect information about people’s personalities, and prior validation of these items and their relation to actual personality traits.

We collected new matrices of face (n = 496), familiar person knowledge (n = 478), and social group trait space (n = 489) using distinct trait descriptions between each task (see

Methods). The data of Study 1 were used for the conceptual similarity matrix. Consistent with our hypothesis, we again observed a significant correlation between the conceptual trait space matrix and the three social perceptual trait space matrices, despite the use of unique items to construct the different matrices (Supplementary Figure 2; conceptual matrix predicts: face trait space matrix, Spearman $\rho(103) = 0.575$, $\rho^2(103) = 0.331$, $p < 0.0001$; 95% CI = [0.431 , 0.691]; familiar person trait space matrix, Spearman $\rho(103) = 0.576$, $\rho^2(103) = 0.332$, $p < 0.0001$; 95% CI = [0.432 , 0.691]; social group trait space matrix, Spearman $\rho(103) = 0.574$, $\rho^2(103) = 0.329$, $p < 0.0001$; 95% CI = [0.430 , 0.690]). Similar results were obtained when controlling for valence (see Supplementary Results). These findings again suggest that trait conceptual associations and inferences are correlated across domains of social cognition in a similar fashion, suggesting that domain-general conceptual associations may be applied across each domain, resulting in a common trait space. The results also suggest that the commonality in trait space is due to beliefs about personality traits as concepts used to predict meaningful social behavior, rather than a mere artifact of semantic relatedness among trait terms.

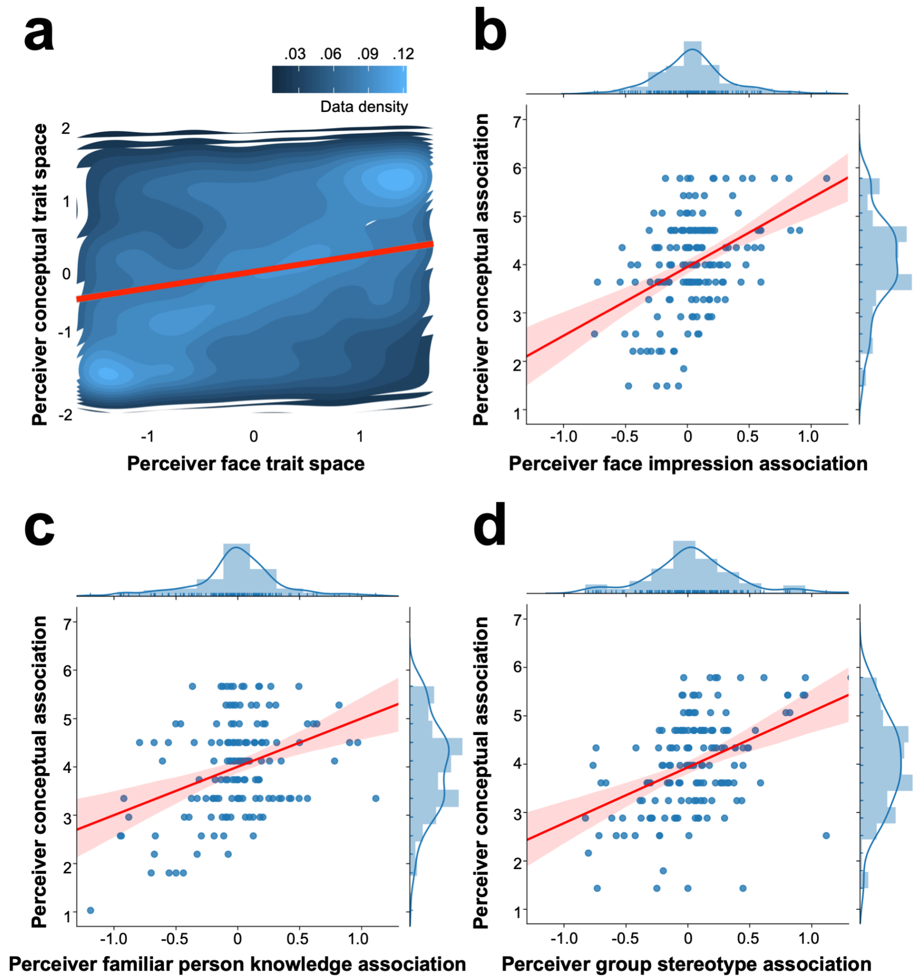


Figure 3. **Individual differences in conceptual knowledge predict social perception**. Study 3 tested whether the subjective conceptual trait space of a perceiver uniquely predicts their face trait space (panel a). A linear mixed-effects model was fit to effectively perform RSA clustered per subject (see Methods), and participant subjective conceptual trait space matrices (*y*-axis) predicted their face trait space matrices (*x*-axis), over and above the group-average conceptual trait space matrix (to isolate the effect of subjective associations; estimate of this fixed effect is plotted; $b = 0.145$, $SE = .020$, $t(141.5) = 7.432$, $p < .0001$, 95% CI = [0.11, 0.19]). Each data-point is a trait-pair (28 pairs), unique to each subject ($n = 162$; total of 4563 data points; a contour plot is provided due to the quantity of data, where the color lightness of the density function represents the probability of each value given the range of values). In Study 4, we see that, across domains, perceivers who believe two traits are more correlated (*y*-axis) also see those traits more similarly in targets (*x*-axis; face impressions, panel b, $n = 167$, Spearman $\rho(165) = 0.331$, $\rho^2(165) = 0.110$, $p < 0.0001$; 95% CI = [0.189 , 0.460]; familiar person knowledge, panel c, $n = 155$, Spearman $\rho(153) = 0.308$, $\rho^2(153) = 0.095$, $p < 0.0001$; 95% CI = [0.158 , 0.444]; group stereotypes, panel d, $n = 162$, Spearman $\rho(160) = 0.435$, $\rho^2(160) = 0.189$, $p < 0.0001$; 95% CI = [0.301 , 0.552]). Error ribbons display standard error of the estimate, and data points are each participant per study. While Pearson correlations are plotted for ease of interpretation, statistical analyses were of rank ordered data points. These results suggest that subjects' idiosyncratic conceptual knowledge and trait inferences are related.

Study 3

While Studies 1 and 2 provide initial evidence that the structure of trait inferences reflect that of conceptual associations, these high-level assessments only ask whether, on average across perceivers, traits more correlated in conceptual associations are more correlated in trait inferences. An important component of our theoretical perspective is that conceptual trait associations may shift initial trait inferences, which would entail that variance between perceivers' conceptual knowledge should uniquely shape their idiosyncratic trait inferences across domains. People who believe two traits are more or less correlated (e.g., 'kind people are/not intelligent') should make more or less tethered inferences of those two traits (e.g., 'kind faces and groups are/not intelligent').

In Study 3 (n = 162), we extended the methodology of Studies 1 and 2. Focusing on face impressions, we collected both conceptual and face trait space matrices per subject along 8 traits ('adventurous', 'assertive', 'cautious', 'depressed', 'emotional', 'friendly', 'self-disciplined', 'trustworthy'). We performed RSA within a linear mixed-effects model (see Methods), predicting participants' face trait space matrices via their subjective conceptual trait space matrices. Importantly, we allowed for random effects of subject and controlled for the group-average conceptual trait space matrix, therefore testing whether there is a unique contribution of subjective conceptual knowledge to face impressions. Subjective conceptual trait space significantly predicted subjective face trait space over and above group-average conceptual trait-space (Fig. 3a; b = 0.145, SE = .020, t(141.5) = 7.432, p < .0001, 95% CI = [0.11, 0.19]; similar results were obtained when controlling for valence; see Supplementary Results).

Study 4

To further explore the role of perceiver's idiosyncratic conceptual structure and individual differences, as well as survey each domain of trait inferences, in Study 4 we tested whether individual differences in conceptual trait associations correspond to individual differences in specific trait inferences in each domain: face impressions (n = 167), familiar person knowledge (n = 155), and social groups stereotypes (n = 162). In this task, each participant first rated target stimuli along a pair of two randomly assigned traits, then reported their conceptual association between the assigned trait-pair. We then tested whether individual differences in conceptual associations correlated with individual differences in trait inference associations. In support of our account, we found a consistent relationship between perceiver conceptual and trait inference associations. The more perceivers conceptually associated trait-pairs the more they saw those traits similarly in targets (Fig. 3b,c,d; conceptual associations correlate with: face impressions, Spearman $\rho(165) = 0.331$, $\rho^2(165) = 0.110$, $p < 0.0001$; 95% CI = [0.189 , 0.460]; familiar person knowledge, Spearman $\rho(153) = 0.308$, $\rho^2(153) = 0.095$, $p < 0.0001$; 95% CI = [0.158 , 0.444]; and social group stereotypes, Spearman $\rho(160) = 0.435$, $\rho^2(160) = 0.189$, $p < 0.0001$; 95% CI = [0.301 , 0.552]). In addition to Study 3, these results demonstrate that perceivers' subjective trait inferences reflect their unique conceptual associations. Importantly, these results suggest a common trait space is observed within perceivers in line with their own subjective conceptual knowledge, and a common yet divergent structure of trait space between perceivers may emerge to the extent perceivers share or diverge in their conceptual trait knowledge.

Study 5

A key premise of our perspective is that conceptual associations between traits are used in the trait inference process, shaping their initial formation and consequently their intercorrelations, from which emerges a conceptually bound trait space across social perception. So far, while we have found evidence for the relationship between conceptual associations and trait inferences, this evidence has been correlational in nature. In Study 5 we manipulated perceiver conceptual associations to more directly examine their directional influence on trait inferences. Participants (*n* = 141) were randomly assigned to one of two between-subjects conditions, in which they were led to believe two traits were either positively or negatively correlated. At the beginning of the study, participants were randomly allocated a trait-pair from the 6 unique pairings of 'friendly', 'depressed', 'intellectual'. To manipulate the direction of conceptual associations, participants read a faux science article about personality, which described research finding that the two traits assigned to that participant were either positively or negatively correlated (see Methods). Participants then completed a face rating task.

Our analysis tested whether the associations of participants' face impressions were affected by their assigned conceptual association. For instance, we predicted that participants led to believe 'friendly' people are more vs. less often 'depressed' would rate friendly appearing faces as more vs. less 'depressed'. As repeated face ratings were nested within participant, we examined our hypothesis in a multilevel model. We regressed average ratings of the faces along one trait dimension (e.g., 'friendly'; average ratings taken from independent raters) on our participants' ratings of the faces along the other trait dimension (e.g., 'intellectual'), their assigned association condition, and interaction of these two variables (see Methods and Supplementary Methods for details). Participant ratings were group-centered. In the model, intercepts were random and slopes were fixed. Consistent with our hypothesis, the strength of association between the two trait dimensions varied by association condition (Fig. 4; $B = 0.023$, $SE = .003$, $p < .0001$, 95% CI = [0.017, 0.030]). Simple slopes revealed a more negative trait-pair association in the negative condition (*B* = -0.084, *SE* =.005, $p < .0001$, 95% CI = [-0.094, -0.074]) than the positive condition (*B* = -0.037, *SE* =.003, $p < .0001$, 95% CI = [-0.047, -0.029]). Note that in both conditions, the trait-pair association should reflect not only the effect of the manipulation but also priors or a 'baseline' level of association between the particular traits. Thus, although the regression coefficient is negative in both conditions, what is critical is whether its magnitude differs across conditions. These results provide some causal evidence in support of our theoretical account that conceptual trait associations structure initial inferences and their correlations, from which trait space emerges.

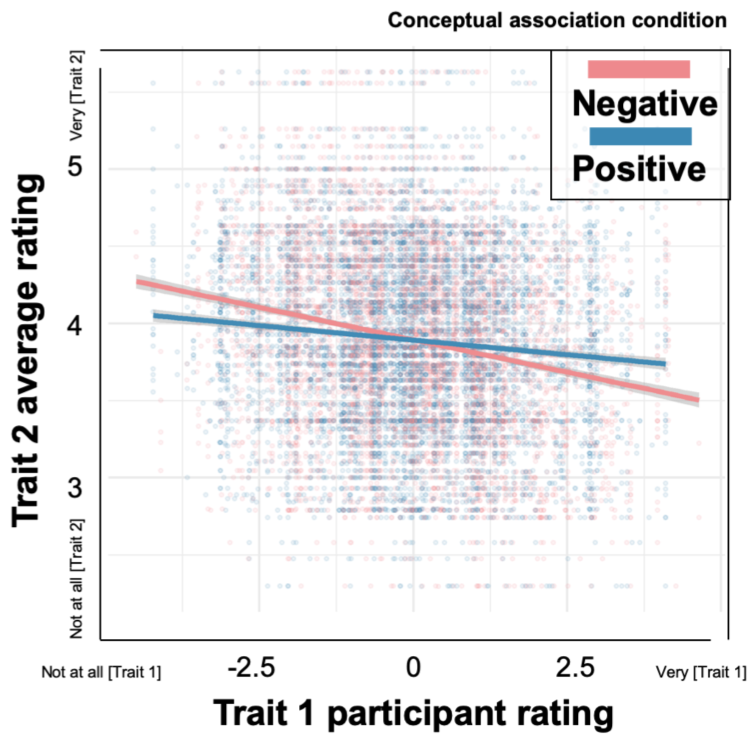


Figure 4. **Conceptual associations directly shape trait inferences and space.** In Study 5, we manipulated perceiver conceptual associations with faux science articles, and found that perceivers in the negative conceptual association condition see traits less similarly in targets compared to participants in the positive association condition (panel d, n = 141, 90 faces rated, with 12690 data points plotted; B = 0.023, SE =.003, p < .0001, 95% CI = [0.017, 0.030]. Red vs. Blue data points and lines distinguish the negative vs. positive association conditions. Error ribbons display standard error of the estimates. Participant ratings of faces (each data point is a face) along one trait (x-axis; 'Trait 1') correlated with average ratings of the face (from Study 1) along another trait (y-axis; 'Trait 2') more negatively in the negative association (B = -0.084, SE =.005, p < .0001, 95% CI = [-0.094, -0.074]) than positive association condition (B = -0.037, SE =.003, p < .0001, 95% CI = [-0.047, -0.029]). (Note that in both conditions, priors likely set baseline associations between traits and tether the manipulation down. Therefore, the magnitude and direction of a slope is not relevant to the effects. What is critical is whether its magnitude differs between conditions.) These findings together demonstrate that conceptual associations may directly influence the initial trait inference process, from which the structure of trait space may emerge.

Study 6

While conceptual associations may guide trait inferences, it is also certainly plausible that this influence is bidirectional: the origins of conceptual trait associations may derive from inferences about the social world. Study 6 provided an initial test of this possibility. Rather than manipulate a conceptual trait association and measure its effect on the correlation of face judgments (as in Study 5), here we test whether the reverse also holds true. In Study 6, participants learned that two traits ('friendly' and 'cautious') were correlated positively or negatively depending on their between-subjects condition. Participants were tasked with assessing the 'cautiousness' of different individuals via their faces. After they rated each face, they were given feedback after the trial as to whether they were correct or incorrect regarding the face's 'cautiousness'. These faces varied in how friendly they appeared. Friendliness judgments are highly consistent across perceivers[22], and we used faces rated low vs. high in perceived friendliness from Study 1 for the present study. In the 'positive association' condition, faces above average in how 'friendly' they looked were labeled as 'more cautious' and faces below average in how 'friendly' they looked were labeled as 'less cautious'. In the 'negative association' condition this pattern was reversed. Therefore, as participants judged the 'cautiousness' of different faces, they received feedback that reinforced either the positive or negative association of targets' 'cautiousness' with the targets' 'friendliness'. Afterward, participants reported their conceptual associations between the two traits as in prior studies.

As predicted, we found that perceivers manipulated to believe friendliness and cautiousness were positively associated conceptually associated the two traits to a stronger degree ($M = 4.508$, $SD = 1.593$) than perceivers led to believe the traits were negatively associated ($M = 4.019$, $SD = 1.416$; mean difference = 0.4888, independent t-test, $t(144) = 2.127$,

$p = .035$, $r^2 = .03$, mean difference 95% CI = [0.035, 0.943]; Fig. 5a). These findings support the hypothesis that perceivers not only apply conceptual knowledge to social perception, but also learn about trait concepts from social perception. However, an important limitation of these experiments is the strong priors that individuals hold through a lifetime of learning that would precede this experiment. More critically, while these findings demonstrate learning of conceptual associations are possible, the results are agnostic to the actual source of the associations that perceivers bring to the table in trait inference (such as those found in Studies 1 – 4).

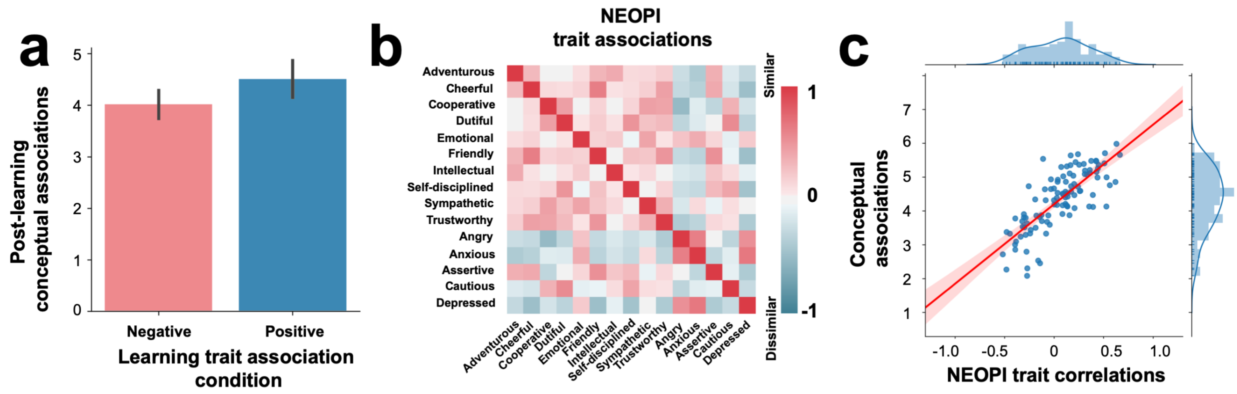


Figure 5. Social perception and trait inferences influence conceptual trait space. In Studies 6 and 7, we tested whether the relationship between conceptual knowledge and trait inferences is bidirectional. In Study 6 (panel a, n = 146), we found conceptual associations between two traits ('cautious', 'friendly'; y-axis) were stronger for participants assigned to observe those two traits positively correlating in target faces (M = 4.508, SD = 1.593), compared to participants assigned to perceive their negative correlation (M = 4.019, SD = 1.416; mean difference = 0.4888, independent t-test, $t(144) = 2.127$, $p = .035$, $r^2 = .03$, mean difference 95% CI = [0.035, 0.943]). Mean (bar height) and standard error (error bars) of participants' conceptual associations are plotted (negative association condition in pink, positive in blue). In Study 7, we test whether perceivers' conceptual knowledge is learned to some extent from the actual structure of human personality. We collected a trait space matrix of actual personality trait correlations of those traits used in prior studies (via the NEOPI; panel b, n = 307,313; personality trait correlations plotted from negative/blue to positive/red). RSA found the NEOPI trait space matrix and conceptual trait space matrices explain a sizeable proportion of variance in one another (panel c, 105 trait pairs as data points, Spearman $\rho(103) = 0.77$, $\rho^2(103) = 0.60$, $p < 0.0001$; 95% CI = [0.684 , 0.841]). Error ribbons display standard error of the estimate, and there are 105 trait-pairs as data points per panel. While Pearson correlations are plotted for ease of interpretation, statistical analyses were of rank ordered data points. These findings suggest conceptual trait space is also shaped through social perception, and one potential source is direct observational or indirect social learning of the actual correlation of personality traits in others.

Study 7

One of the candidate sources of conceptual trait space is learning of the actual structure of others' personality traits. Much like conceptual trait space, people's actual personality traits are highly correlated along a relatively small set of dimensions (e.g., the 'Big Five' factors of personality[23,24]). If actual personality traits were in fact correlated, a simple strategy to optimize trait inference for perceivers would be to learn this structure and make predictions accordingly. Not all of our personality traits are worn on our sleeve[25], so perceivers may take trait information at hand to surmise the whole of a target. For example, perceivers may use 'talkativeness', a more visible trait, to infer a target's 'anxiety', a less visible trait, based on their conceptual association between 'talkative' and 'anxious'[25]. If perceivers learn the actual structure of personality, traits they believe are more similar conceptually would also be more similar in actual human personality structure. This of course would be only one amongst many candidate sources for trait knowledge[26].

To test this possibility, we compared conceptual trait space (as measured in Study 1) to an estimate of actual personality trait space via the NEO personality inventory (henceforth referred to as 'NEOPI trait space'[19,24]). The NEOPI is a canonical and empirically validated model of personality structure, ideal for the current research as participants whose personality traits are measured do not explicitly evaluate whether they possess the traits of 'trustworthiness' or 'anxiety'. This greatly reduces the potential confound that our NEOPI trait space matrix could reflect perceivers' social cognitive trait spaces merely due to semantic similarities in measurement (e.g., reporting of 'warmth' trait in self and in others). The Big-5 factor model of personality is composed of a larger set of personality traits underlying each of the Big five factors, known as its 'facets', which were the trait adjectives used in our research above taken from the facet subscales of the NEOPI[19,27]. Therefore, we were able to calculate a NEOPI trait space matrix comparable the social cognition trait space matrices used in Studies 1 and 2, as the same 15 traits are measured all domains. The NEOPI trait space (Fig. 5b) was computed via data acquired from a large open source dataset (n = 307,313 participants; retrieved from https://osf.io/tbmh5[27]). This allowed us to effectively test whether trait-pairs associated in conceptual knowledge are also associated in a ground-truth model of personality.

Strikingly, perceiver social conceptual knowledge (via Study 1) closely tracked the NEOPI trait structure, where trait-pairs perceivers' conceptually relate are also more correlated in personality as measured in the NEOPI (Fig. 5c; Spearman $\rho(103) = 0.77$, $\rho^2(103) = 0.60$, $p <$ 0.0001; 95% CI = [0.684 , 0.841]). Supplementary analyses confirmed that social perceptual trait spaces also strongly resembled NEOPI trait space, as would be expected through transitivity given our hypothesis that social perceptual trait spaces reflect conceptual trait space (see Supplementary Figure 3 and Supplementary Results.) These findings show that perceiver trait conceptual knowledge reflects actual personality structure, suggesting the possibility that its structure may be learned through some mechanism, such as cultural transmission or accurate observation.

Discussion

Taken together, our results broadly demonstrate that conceptual trait knowledge shapes trait inferences across distinct domains of social perception, including face impressions, familiar person knowledge, and group stereotypes. The similarity structures of social perceptual trait inferences were all highly correlated with that of conceptual trait space (Studies 1 & 2). Participants' idiosyncratic conceptual knowledge was reflected in their inferences and social perceptual trait spaces (Studies 3 & 4), and manipulation of perceiver conceptual associations influenced trait inferences accordingly (Study 5). To probe the source of trait concept knowledge, we found evidence suggesting that conceptual knowledge may be learned through social perception, demonstrating the bidirectional nature of this process, through direct observation (Study 6), or learning about the actual structure of personality traits (Study 7).

The findings provide quantitative evidence for a common trait space across social cognition[2], which may emerge as trait inferences are similarly shaped by learned conceptual trait space across the many domains of social perception. A prominent perspective is that a common trait space arises due to the adaptive utility of its core dimensions – namely that, across social domains, perceivers track those traits significant to our function and survival (intentions and capabilities; e.g., 'competence' and 'warmth'[2,6]). Evidence suggests this is the case, as traits with adaptive utility play a central role in dimensions of social cognition. Yet there is much additional covariation in the true expanse of trait space that is less easily explained by this functional perspective, such as the perceived relationships between humor, sociability, risk aversion, or neuroticism. The findings reported here support a parsimonious explanation by a more proximal mechanism to perceptions, that perceivers' conceptual knowledge about how traits correlate in others shape how correlated they are in social perceptions regardless of their domain[10,13]. This perspective provides a unifying framework through which we may understand trait space as part of a dynamic cognitive process, from which we may generate broad and general hypotheses about social perception based on context-varying models of social conceptual structure. This perspective also fits trait inferences generally, especially outside of face perception, into an emerging picture of the conceptual nature of social perception[28,29].

This flexible account may be indispensable for accommodating emerging findings of dynamic shifts in social cognitive models. Variation in social trait spaces has been increasingly well documented, suggesting trait space may in fact be dynamic rather than fixed in its structure, both shifting in its core dimensions and their relations depending on social factors[18,30-35]. While trait space generally tends to be consistent across perceivers, various perceiver factors (e.g., stereotypes, motivations, emotions) and social context may shape trait space (for review, see[9]), as much of the variance in trait inferences is due to perceiver characteristics[22]. A trait space shaped by perceiver conceptual knowledge could, in theory, underlie these various findings. For instance, competence and warmth inferences come to correlate positively towards liked groups[36] and negatively towards disliked groups or groups with specific stereotypes (e.g., outgroups and women[30,37]). Perhaps conceptual associations between personality traits vary across these contexts in systematic ways. Future research should investigate how conceptual associations shift across social contexts, and whether these shifts are reliably reflected in different social trait spaces.

An important question concerns the origins of perceivers' conceptual trait associations, which we argue may lie at the foundation of a common social trait space. Human personality traits are in fact intercorrelated[23,24], and this similarity structure is tied to patterns of behavior[38]. Thus, it is possible people may come to learn actual personality structure to predict others' behavior[10]. Here we found that conceptual trait space reflects that of actual personality traits. Prior research has found similar associations[13], and that perceivers can use accurate knowledge of one personality trait to accurately predict other traits of which they are not explicitly informed[39]. While our findings suggest perceivers learn actual personality structure, this is an area ripe for future research. Trait knowledge may be acquired through direct observation, such as social and statistical learning of the social environment[40], or indirect sources, such as cultural learning and gossip[41]. Such knowledge is also likely shaped and biased by the host of processes and biases long known to influence trait inferences[26]. One interesting question is relative degree semantic knowledge compared to cognitive biases contribute to trait space structure. Furthermore, our findings do not speak to which domains of social perception provide information about actual personality. We would speculate different domains must contribute differently. For instance, perception of more or less familiar individuals may provide more or less signal towards actual trait correlations[25,42], yet perceived trait correlations in faces and stereotypes may suffer from limited signal as these sources are often biased[43-45,], cf.[46], and thus contribute less significantly to such accurate trait correlation learning. This line of research may therefore be of interest to the accuracy literature more broadly[47], and join other recent findings exploring accuracy through the perspective of trait space models[15]. Future research should quantify the contributions of different information sources to the development of conceptual trait knowledge.

It is crucial to note that, although trait space structure may be learned from actual personality structure, this should not imply perceivers' persistent accuracy in trait inferences themselves across domains, especially when perceivers begin with inaccurate and biased inferences (e.g., in case of face impressions[44]). Rather, an accurately learned trait space may just as often lead to broad inaccuracies. Humans often begin with inaccurate and biased trait inferences. When initial inferences are inaccurate, other trait inferences made through what are accurate associations may increase in likely inaccuracy. For instance, if friendliness and sociability personality traits are in fact correlated, and perceivers understand this, an erroneous 'unfriendly' inference of a friendly target would lead to an 'unsociable' inference, although the target is more likely to be sociable. Thus, an accurate trait space structure is easily misapplied by inaccurate inference content, and the structure of trait space can be an accurate reflection of reality while the content of inferences is far from it. It will be important for future work to identify when trait space may lead to accuracy or error in judgment.

We believe our perspective and findings here suggest a reorienting in the study of trait space is needed. These and other recent findings[14,17] suggest trait space as a key process in forming initial perceptions[7], in which trait space as measured in the context of target evaluations is merely an emergent property of this process. Some of the most interesting questions may be how and when conceptual trait space is used, for better or worse, and what unique predictions this framework affords models of social perception. One salient prediction, much like in the case of individuation in stereotyping[48], is that trait space is most used when other trait information is scarce. This is akin to saying perceivers may have trait stereotypes, or make further trait generalizations based upon those they initially infer. In recent years, a resurgence in the study of social perceptual dimensions has occurred, with scientific interest in what dimensions best describe trait or mental state space[9,15]. Should trait space be dynamic and context-dependent, attempts to identify and refine any 'true' universal dimensions may be misguided, as trait space is destined to vary when its conceptual basis and its application does. Future research might benefit from explicating the precise role and shape of trait space in the context at hand, from which we may develop nuanced models that predict the structure of trait inferences in particular domains.

There are several limitations of this work. First, although we manipulate conceptual associations or face trait covariations in Studies 5 and 6, more thorough designs should be developed to test the bidirectional and mutually causal relationship between conceptual trait associations and social perception Another limitation is the use of verbal stimuli to measure trait space, as it may influence participant responses to conform across these tasks, especially between actual and perceived trait spaces (for a review of related research, see[10]). Lastly, Studies 6 and 7 are demonstrations of possible sources of information shaping conceptual associations and should not be taken to present an exhaustive argument for the origins of conceptual social knowledge.

In short, the present research provides evidence of a common trait space across social cognition, structured by perceivers' learned conceptions of how personality traits correlate. This account of trait space not only explains its homogeneity across social cognition, but highlights that trait space may be dynamic rather than fixed in its architecture to the extent perceiver conceptual knowledge about personality shifts due to myriad social and contextual factors. We hope this work provides a parsimonious framework to understand trait space, and importantly, allows us to move beyond its measurement to questions of its foundational role in social perception.

Methods

All studies here conducted comply with ethical regulations for research on human subjects and all participants gave informed consent, as approved by the University Committee on Activities Involving Human Subjects at New York University. Subjects were financially compensated $0.10 USD per minute for their participation. Statistical tests are two-tailed. Data distributions were assumed to be normal but this was not formally tested. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications [17,29]. Randomization was applied where possible in all studies, and is described explicitly in each study methods section. Data collection and analysis were not performed blind to the conditions of the experiments.

All data, stimuli names, and data preparation and analysis code are available on the Open Science Framework (https://osf.io/2uzsx/). Analyses were performed in Python and R. Additional details on task instructions and approach to data preparation and analysis are provided in the Supplementary Methods section of the Supplementary Information.

Study 1

Participants. We aimed to recruit ample raters (participants) to acquire stable and reliable estimates of the trait ratings per each exemplar. Our target sample was 30 participants per trait rated in each of the social perception rating tasks below: face, familiar person, and social group trait tasks, as trait ratings across traits stabilize at approximately this number of participant raters[49]. Across traits and tasks, this totaled a target sample of $n = 450$ per social perception model. For the conceptual trait space model, involving conceptual ratings of traits with other traits, we based target sample size on prior work estimating a similar model[29], seeking a target sample of $n = 100$.

Conceptual trait task. We collected conceptual trait association data from 116 subjects via Amazon Mechanical Turk (demographic data missing for 1 subject; all U.S. residents; all primary English-speakers; $M_{\text{age}} = 35.4$ years, $SD_{\text{age}} = 10.5$ years; 58 Female, 55 Male, 2 other; 113 White, 2 other).

Face trait task. We collected face impression data from 484 subjects via Amazon Mechanical Turk (demographic data missing for 2 subjects; all U.S. residents; all primary English-speakers; $M_{\text{age}} = 35.5$ years, $SD_{\text{age}} = 12.3$ years; 281 Female, 199 Male, 2 other; 372 White, 44 Black, 31 Asian, 37 other). Participants were randomly assigned to evaluate one personality trait in all face stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (32 participants per trait condition on average).

Familiar person trait task. We collected familiar person knowledge data from 503 subjects via Amazon Mechanical Turk (demographic data missing for 4 subjects; all U.S. residents; all primary English-speakers; M_{age} = 30.7 years, SD_{age} = 7.1 years; 308 Female, 175 Male, 16 other; 368 White, 44 Black, 42 Asian, 49 other). Participants were randomly assigned to evaluate one personality trait in all familiar person stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (≈34 participants per trait condition).

Group trait task. We collected group stereotype content data from 488 subjects via Amazon Mechanical Turk (demographic data missing for 3 subjects; all U.S. residents; all primary English-speakers; M_{age} = 30.4 years, SD_{age} = 6.9 years; 297 Female, 183 Male, 5 other; 368 White, 44 Black, 39 Asian, 37 other). Participants were randomly assigned to evaluate one personality trait in all group stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (≈33 participants per trait condition).

Valence task. We collected valence ratings of each personality trait adjective used in the above tasks from 69 subjects via Amazon Mechanical Turk (n = 69; M_{age} = 31.4 years, SD_{age} = 6.6 years; 28 Female, 40 Male, 1 other; 52 White, 11 Black, 4 Asian, 2 other).

Stimuli.

Personality trait stimuli. We chose personality trait stimuli which corresponded with many of the facet subscales of the NEOPI[19,27]. We chose 15 facet subscale traits, including three from each of the ‘Big 5’ personality factors to maintain a balance with the comparison of actual personality trait space. These were sub-traits of the primary five-factors: ‘Agreeableness’, ‘Conscientiousness’, ‘Extroversion’, ‘Neuroticism’, and ‘Openness’. The three chosen per primary factor were selected to most easily translate into adjectives participants could engage comfortably in each task. These traits included: ‘adventurous', ‘angry’, ‘anxious’, ‘assertive’, ‘cautious’, ‘cheerful’, ‘cooperative’, ‘depressed’, ‘dutiful’, ‘emotional’, ‘friendly’, ‘intellectual’, 'self-disciplined', 'sympathetic', and 'trustworthy’.

Face stimuli. All stimuli were taken from the Chicago Face Database[50]. Face stimuli included 90 portrait photographs of young White male individuals with neutral facial expressions. Exact stimulus identification numbers are provided in the OSF page (https://osf.io/2uzsx/). Example stimuli are presented in Fig. 1b.

Familiar person stimuli. All familiar person stimuli were chosen from recent work that used data-driven methods to identify individuals highest in familiarity to a similar online sample demographic, and maximize diversity in traits of the stimuli to guarantee a wide and generalizable sampling of trait space[51]. We used all 60 familiar person stimuli identified in Thornton and Mitchell [51]. Stimuli are provided in the OSF page (https://osf.io/2uzsx/). Example stimuli are presented in Fig. 1c.

Group stimuli. In order to obtain a diverse set of social group stimuli, we chose the 80 most frequently named social groups in the U.S., as named in recent work by an online participant demographic similar to our own[33]. Stimuli are provided in the OSF page (https://osf.io/2uzsx/). Example stimuli are presented in Fig. 1d.

Protocol.

Conceptual trait task. Participants were informed they would partake in a study on how different personality traits correlate in the world. After several examples and practice trials, participants began the task. Each trial item asked, “Given that an individual possesses one trait, how likely is it that they possess the other?”, then presented the two trait stimuli for that trial separated by a hyphen (e.g., ‘friendly – self-disciplined’). Participants evaluated the conceptual relationship of each trait-pair in the 15 trait stimuli (1 – 7 Likert-type scale, 1 - “Not at all likely”

— 7 – “Very likely”), presented in both order given the wording of the item question (e.g., ‘friendly – self-disciplined’ and ‘self-disciplined – friendly’). Therefore, there were a total of 210 trials for each participant (total possible permutations from all pairs of 15 trait stimuli). Participants then completed a general demographics survey.

Face trait task. Participants were informed they would partake in a study examining how people perceive others. Each participant was randomly assigned to evaluate only one of the 15 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 – ‘Not at all friendly’, 4 – ‘neutral’, 7 – ‘Very friendly’). Following the face trait rating task, participants completed a general demographics survey.

Familiar person trait task. Participants were informed they would partake in a study examining how people perceive famous people. Each participant was randomly assigned to evaluate only one of the 15 personality trait stimuli in familiar person stimuli. In the task, participants rated each of the 60 familiar person stimuli on the personality trait they were assigned (1 – 7 Likert-type scale; e.g., 1 – ‘Not at all friendly’, 4 – ‘neutral’, 7 – ‘Very friendly’). Following the familiar person trait rating task, participants completed a general demographics survey.

Group trait task. Participants were informed they would partake in a study examining common societal inferences, rather than their own, towards common social groups. Instructions were intended to reduce the influence of social desirability on responses[5,33]. Each participant was randomly assigned to evaluate only one of the 15 personality trait stimuli in the social group stimuli. In the task, participants rated each of the 80 social group stimuli on the personality trait they were assigned (1 – 7 Likert-type scale; e.g., 1 – “Not at all friendly” — 7 – “Very friendly”). Following the social group trait rating task, participants completed a general demographics survey.

Valence task. Participants were instructed to rate personality traits on their valence, or how negative to positive the trait is. Participants then rated one stimulus at a time (1 – 7 Likert-type scale; e.g., 1 – ‘Very negative’ — 7 – ‘Very positive’). Trials were randomized per subject. Participants responded to a basic demographics survey after the task.

Study 2

Participants.

Face trait task. We collected face impression data from 496 subjects via Amazon Mechanical Turk (all U.S. residents; all primary English-speakers; M_{age} = 30.3 years, SD_{age} = 6.3 years; 257 Female, 237 Male, 2 other; 320 White, 101 Asian, 30 Black, 45 other). Participants were randomly assigned to evaluate one personality trait in all face stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (≈33 participants per trait condition on average).

Familiar person trait task. We collected familiar person knowledge data from 478 subjects via Amazon Mechanical Turk (demographic data missing for 2 subjects; all U.S. residents; all primary English-speakers; M_{age} = 29.8 years, SD_{age} = 6.3 years; 239 Female, 237 Male, 2 other; 309 White, 89 Asian, 40 Black, 40 other). Participants were randomly assigned to evaluate one personality trait in all familiar person stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (≈32 participants per trait condition).

Group trait task. We collected group stereotype content data from 489 subjects via Amazon Mechanical Turk (all U.S. residents; all primary English-speakers; M_{age} = 30.4 years, SD_{age} = 6.7 years; 263 Female, 223 Male, 3 other; 315 White, 89 Asian, 43 Black, 42 other).

Participants were randomly assigned to evaluate one personality trait in all group stimuli, and were therefore divided roughly equally between all 15 personality trait conditions (≈33 participants per trait condition).

Stimuli.

Personality trait stimuli. We replaced each of the original 15 trait items from Study 1 with items that asked about the likely behavior of the target. A different item was used in each domain (i.e., item A would only be used in the face task, item B in the familiar person task, and item C in the social group task). We chose the new trait stimuli to replace trait terms with from the NEOPI facet items. In the NEOPI, to measure personality, participants are not asked directly whether they are 'kind', but asked multiple items that describe behavioral tendencies that have been found to relate to kindness. Given the long history of validation of these items and their correspondence to behaviors that relate to underlying personality traits, we chose our new trait stimuli from these items. Specifically, for each of the 15 traits in our similarity matrices, we chose 3 NEOPI items, so that one could be used in each of the three social perception tasks (face, familiar person, social group). Specific items may be viewed via the OSF (https://osf.io/2uzsx/).

Target stimuli. For face, familiar person, and social group target stimuli, Study 2 used the same stimuli as Study 1.

Protocol. Study 2 used an identical task design for each of the three tasks as Study 1, where only the items were replaced (see above).

Study 3

Participants.

All data and tasks were performed by and within each participant. We collected face impression and conceptual association data from 162 subjects via Amazon Mechanical Turk (original *n* = 168; 6 subjects dropped due to failure to follow task instructions; all United States residents; all primary English-speakers; M_{age} = 31.9 years, SD_{age} = 5.9 years; 54 Female, 108 Male; 113 White, 33 Black, 9 Asian, 7 other).

Stimuli. As each subject completed multiple face ratings tasks, and a conceptual association task (compared to single trait inference tasks in Studies 1 and 2), a subset of trait adjective and face stimuli were used in Study 3 in consideration of time constraints and participant fatigue.

Personality trait stimuli. We chose a subset set of trait stimuli from those used in Studies 1 and. Trait stimuli included: ‘adventurous’, ‘assertive’, ‘cautious’, ‘depressed’, ‘emotional’, ‘friendly’, ‘self-disciplined’, and ‘trustworthy’. We used each trait in its own single block face rating task, and each pairwise combination of these traits in the conceptual association task.

Face stimuli. For face stimuli, each participant was assigned to a random subset of 25 stimuli from the face stimulus set used in Studies 1 and 2.

Protocol. Participants first completed one block of face ratings for each personality trait stimulus, with the trait block order randomized (one trait within each block, for a total of 8 blocks). The same 25 face stimuli were rated within each block. Following the series of face rating tasks, participants completed a conceptual association task in which they rated their pairwise conceptual association of each trait-pair presented in random order. Each task otherwise had an identical design to that of Study 1. Following completion of the two task sets, participants completed a standard demographics survey.

Study 4

Participants.

Face trait task. We collected face impression data from 167 subjects via Amazon Mechanical Turk (original n = 174; 5 subjects dropped due to task incompletion; 2 subjects dropped due to failure to follow task instructions; all United States residents; all primary English-speakers; M_{age} = 31.44 years, SD_{age} = 5.50 years; 102 Female, 64 Male, 1 decline; 128 White, 23 Black, 5 Asian, 11 other).

Familiar person trait task. We collected familiar person knowledge data from 155 subjects via Amazon Mechanical Turk (original n = 167; 9 subjects dropped due to task incompletion; 3 subjects dropped due to failure to follow task instructions; all United States residents; all primary English-speakers; M_{age} = 32.34 years, SD_{age} = 6.52 years; 70 Female, 82 Male, 3 decline; 120 White, 20 Black, 6 Asian, 19 other).

Social group trait task. We collected group stereotype content data from 162 subjects via Amazon Mechanical Turk (original n = 168; 6 subjects dropped due to task incompletion; all United States residents; all primary English-speakers; M_{age} = 31.45 years, SD_{age} = 5.53 years; 72 Female, 90 Male; 126 White, 20 Black, 8 Asian, 8 other).

Stimuli.

Personality trait stimuli. We chose a diverse set of trait stimuli somewhat deviating from those in Study 1 to assess generalizability. Trait stimuli included: ‘creative, ‘dishonest’, ‘friendly, ‘intelligent, ‘sociable, and ‘stubborn’. We used all pairwise combinations of these trait pairs (for a total of 15 unique possible trait-pairs). Participants were randomly assigned to one of the 15 total trait-pair combinations.

Target stimuli. For face, familiar person, and social group target stimuli, Study 4 used the same stimuli as Study 1.

Protocol. Both social perception trait and conceptual trait tasks were largely identical in design within themselves to those in previous studies (see Study 1 methods). A major distinction is that in this study, each participant both provided target trait and conceptual trait data. Each participant was randomly assigned to one of 15 trait-pairs (the unique combinations of 6 trait stimuli: ‘creative, ‘dishonest’, ‘friendly, ‘intelligent, ‘sociable, and ‘stubborn’). First, participants evaluated all stimuli on both assigned traits (either face, familiar person, or group stimuli depending on the task). They evaluated all stimuli on one trait first, followed by the other, the order of which trait came first was randomized. The order of which trait was first evaluated was randomly determined per subject. In total, participants therefore completed: 180 trials of face impressions, 120 trials of familiar person impressions, 160 trials of group inferences. From this data, we were able to measure the correlation of inferences 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 (as compared to the many trait-pairs in Study 1), there were only 2 trials in the conceptual trait task, randomly ordered. Following these tasks, participants completed a general demographics survey.

Study 5

Participants. We collected face impression data from 141 subjects via Amazon Mechanical Turk (original n = 192; 51 subjects dropped due to insensitivity to experiment manipulation; all United States residents; all primary English-speakers; M_{age} = 32.69 years, SD_{age} = 6.44 years; 64 Female, 77 Male; 102 White, 18 Black, 9 Asian, 12 other).

Stimuli.

Personality trait stimuli. We chose 3 trait terms from the facets of the Big Five factors of personality, corresponding to the ‘Agreeableness’, ‘Neuroticism, and ‘Openness’ factors:

'friendly', 'depressed', and 'intellectual' (for 3 combinations of trait-pairs). These traits were chosen given their correspondence to both relatively independent trait concepts, and to traditional dimensions of social perception ('friendly' to 'warmth', 'intellectual' to 'competence')[2] and core personality traits ('friendly' to 'agreeable', 'intellectual' to 'openness', and 'depressed' to 'neuroticism')[19] . Further, we chose traits whose conceptual associations could be realistically manipulated in perceivers (e.g., given likely strong priors for associations of traits that load along the same factors, such as 'warmth' and 'sociability').

Trait association manipulation article. In order to manipulate participant conceptual associations between traits, participants read a fake scientific article about the actual correlation of personality traits. Participants read an adapted article from prior research that was successful in manipulating lay theories of gender[52]. The article explained a research study conducted by personality researchers, who find that on average individuals with one personality trait (e.g., friendliness) are more or less likely to have another personality trait (e.g., depression). Each participant was randomly assigned to one trait-pair at the beginning of the study, which was inserted into the article. The manipulation articles are available on the OSF (https://osf.io/2uzsx/).

Face trait stimuli. Study 5 used the same face stimuli as Study 1.

Manipulation check. After the experiment, participants completed a brief questionnaire to assess effectiveness of the manipulation. Modeled from prior research and our own measurement methods[17,52], participants were asked direct questions about their conceptual associations between their assigned trait-pair to assess manipulation effectiveness. We asked participants how likely individuals with the first assigned personality trait are likely to have the second trait assigned to the participant, and vice versa (e.g., 'How likely is a friendly person to be intellectual?', 'How likely is an intellectual person to be intelligent?'; Likert-type scale, 1 — Not at all likely, 2, 3, 4 — Neutral, 5, 6, 7 — Very likely).

Protocol. At the beginning of the study, participants were randomly assigned to one of two 'association direction' conditions, specifying whether the between-subjects manipulation would convince them their trait-pair was negatively or positively correlated (e.g., are 'friendly' people more likely to be 'depressed' or less likely to be 'depressed'). Participants were randomly allocated one of the three trait-pairs. Participants were informed they would take part in a study on how people think of others. In the first part, we manipulated what they think about personality by having them read an article about personality. Once participants began to read the article, we did not allow them to progress past the article for 2 minutes to further encourage their reading and engagement of the article given its length. The article explained research finding the participant's trait-pair (e.g., 'depressed'-'friendly') was negatively or positively correlated, depending on the participant's association direction condition. After reading the article, participants were given a free response form to summarize the article and additionally provide their thoughts on the article and personality generally. Next, we informed participants a new task would begin where they would make personality judgments of others based on only their face. This task and its instructions were identical to that of the face rating tasks in prior studies. Participants rated all 90 face stimuli on one trait to which they were assigned (trait randomly chosen; Likert-type scale, e.g., 1 — Not at all friendly, 4 — Neutral, 7 — Very friendly). Lastly, participants completed the manipulation check, reporting their conceptual association for their assigned trait-pair. Instructions and item design were identical to those used in Study 3.

Study 6

Participants. We collected data from 146 subjects via Amazon Mechanical Turk (*n* = 146; 6 subjects dropped due to task incompletion; all United States residents; all primary English-speakers; M_{age} = 37.2 years, SD_{age} = 12.7 years; 69 Female, 77 Male; 115 White, 15 Black, 9 Asian, 7 other).

Stimuli.

Personality trait stimuli. Participants guessed the ostensible ‘cautiousness’ of face stimuli. We next measured their conceptual association between ‘friendly’ and ‘cautious’. ‘Friendliness’ was chosen due to the spontaneity of ‘friendly’ face impressions[53,54]., and we chose 'cautiousness' because of its relatively low conceptual association with 'friendliness' (where in Study 1 it had the lowest relationship, closest to the 'neutral' response option, with a value of 4.13 on the 1-7 Likert-type scale).

Face stimuli. Face stimuli were a subset of 56 face stimuli from those used in prior experiments. The stimuli were split into two sets based upon their above or below ‘Neutral’ ‘friendly’ ratings from Study 1, allowing us to label responses to the more or less ‘friendly’ faces as ‘more cautious’ or ‘less cautious’ depending on the subjects’ experimental conditions.

Protocol. The task was a two-part task, consisting of the learning phase, in which participants were manipulated to either positively or negatively associate ‘friendliness’ with ‘cautiousness’, followed by the conceptual trait association task, in which they reported their conceptual association between the personality trait stimuli. In the learning phase, participants guessed the ostensible ‘cautiousness’ of the faces, making a two-choice categorization as to whether each face was ‘less cautious’ or ‘more cautious’. Based on their experimental condition, feedback to ‘cautiousness’ judgments of low vs. high ‘friendliness’ faces indicated an incorrect or correct response. These data were not analyzed as the task purpose was manipulation of an association, through feedback in which friendly or unfriendly faces were said to be more or less cautious. Following the feedback phase, they completed a conceptual trait association task identical to that in previous studies (but here it only included the 'friendly' and 'cautious' trait-pair ratings).

Study 7

Data utilized in Study 7 analyses came from the Study 1 conceptual trait task data, and a personality measurement dataset available from prior published research via the Open Science Framework (OSF)[27]. For reporting of methods and data selection from this outside dataset, see the Supplementary Methods.

References

1 Leary, T. Interpersonal diagnosis of personality. *American Journal of Physical Medicine & Rehabilitation* **37**, 331 (1958).

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

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

4 Rosenberg, S., Nelson, C. & Vivekananthan, P. A multidimensional approach to the structure of personality impressions. *Journal of personality and social psychology* **9**, 283 (1968).

5 Fiske, S. T., Cuddy, A. J., Glick, P. & Xu, J. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology* **82**, 878-902 (2002).

6 Cuddy, A. J., Fiske, S. T. & Glick, P. The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology* **92**, 631-648 (2007).

7 Asch, S. E. Forming impressions of personality. *The Journal of Abnormal and Social Psychology* **41**, 258 (1946).

8 Wojciszke, B. Morality and competence in person-and self-perception. *European review of social psychology* **16**, 155-188 (2005).

9 Stolier, R. M., Hehman, E. & Freeman, J. B. A Dynamic Structure of Social Trait Space. *Trends in cognitive sciences* **22**, 197-200, doi:https://doi.org/10.1016/j.tics.2017.12.003 (2018).

10 Schneider, D. J. Implicit personality theory: A review. *Psychological bulletin* **79**, 294 (1973).

11 Peabody, D. & Goldberg, L. R. Some determinants of factor structures from personality-trait descriptors. *Journal of personality and social psychology* **57**, 552 (1989).

12 Digman, J. M. Higher-order factors of the Big Five. *Journal of personality and social psychology* **73**, 1246 (1997).

13 Lay, C. H. & Jackson, D. N. Analysis of the generality of trait-inferential relationships. *Journal of personality and social psychology* **12**, 12 (1969).

14 Thornton, M. A. & Tamir, D. I. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences* **114**, 5982-5987 (2017).

15 Tamir, D. I. & Thornton, M. A. Modeling the Predictive Social Mind. *Trends in cognitive sciences* **22**, 201-212, doi:https://doi.org/10.1016/j.tics.2017.12.005 (2018).

16 Sutherland, C. A. *et al.* Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin* **44**, 521-537 (2018).

17 Stolier, R. M., Hehman, E., Keller, M. D., Walker, M. & Freeman, J. B. The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1807222115 (2018).

18 Xie, S. Y., Flake, J. K. & Hehman, E. Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of personality and social psychology* (2018).

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

20 Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 4, doi:10.3389/neuro.06.004.2008 (2008).

21 Gara, M. A. & Rosenberg, S. Linguistic factors in implicit personality theory. *Journal of personality and social psychology* **41**, 450 (1981).

22 Hehman, E., Sutherland, C. A., Flake, J. K. & Slepian, M. L. The Unique Contributions of Perceiver and Target Characteristics in Person Perception. *Journal of personality and social psychology* **113**, 513 (2017).

23 Goldberg, L. R. The structure of phenotypic personality traits. *American psychologist* **48**, 26 (1993).

24 Digman, J. M. Personality structure: Emergence of the five-factor model. *Annual review of psychology* **41**, 417-440 (1990).

25 Vazire, S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of personality and social psychology* **98**, 281 (2010).

26 Uleman, J. S. & Kressel, L. M. A brief history of theory and research on impression formation. *Oxford handbook of social cognition*, 53-73 (2013).

27 Johnson, J. A. Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality* **51**, 78-89 (2014).

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

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

30 Cuddy, A. J. C. *et al.* Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology* **48**, 1-33, doi:10.1348/014466608X314935 (2009).

31 Weisman, K., Dweck, C. S. & Markman, E. M. Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences* **114**, 11374-11379 (2017).

32 Sutherland, C. A. *et al.* Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition* **127**, 105-118 (2013).

33 Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C. & Alves, H. The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of personality and social psychology* **110**, 675 (2016).

34 Collova, J., Sutherland, C. & Rhodes, G. Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of personality and social psychology* (2019).

35 Yzerbyt, V. Y., Kervyn, N. & Judd, C. M. Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin* **34**, 1110-1123 (2008).

36 Kraft-Todd, G. T. *et al.* Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PloS one* **12**, e0177758 (2017).

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

38 Paunonen, S. V. & Ashton, M. C. Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology* **81**, 524 (2001).

39 Burron, B. F., Carlson, K. A., Getty, G. R. & Jackson, D. N. The effects of informational characteristics on the perception of real and hypothetical target persons. *Psychonomic Science* **23**, 145-147 (1971).

40 Jackson, D. N. A model for inferential accuracy. *Canadian Psychologist/Psychologie Canadienne* **13**, 185 (1972).

41 Baumeister, R. F., Zhang, L. & Vohs, K. D. Gossip as cultural learning. *Review of general psychology* **8**, 111-121 (2004).

42 Paunonen, S. V. Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of personality and social psychology* **56**, 823 (1989).

43 Judd, C. M. & Park, B. Definition and assessment of accuracy in social stereotypes. *Psychological review* **100**, 109 (1993).

44 Todorov, A., Olivola, C. Y., Dotsch, R. & Mende-Siedlecki, P. Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology* **66** (2015).

45 Tskhay, K. O. & Rule, N. O. Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review* **17**, 72-86 (2013).

46 Jussim, L., Crawford, J. T. & Rubinstein, R. S. Stereotype (in) accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science* **24**, 490-497 (2015).

47 Funder, D. C. On the accuracy of personality judgment: a realistic approach. *Psychological review* **102**, 652 (1995).

48 Fiske, S. T. & Neuberg, S. L. A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology* **23**, 1–74 (1990).

49 Hehman, E., Xie, S. Y., Ofosu, E. K. & Nespoli, G. A. Assessing the point at which averages are stable: A tool illustrated in the context of person perception. *Retrieved from https://psyarxiv.com/2n6jq* (2018).

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

51 Thornton, M. A. & Mitchell, J. P. Theories of Person Perception Predict Patterns of Neural Activity During Mentalizing. *Cerebral cortex* **28**, 3505-3520 (2017).

52 Coleman, J. M. & Hong, Y.-Y. Beyond nature and nurture: The influence of lay gender theories on self-stereotyping. *Self and Identity* **7**, 34-53 (2008).

53 Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A. & Hehman, E. A. Amygdala responsivity to high-level social information from unseen faces. *The Journal of Neuroscience* **34**, 10573-10581 (2014).

54 Willis, J. & Todorov, A. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science* **17**, 592-598 (2006).

Data availability statement: Experiment materials information and all experiment de-identified data are publicly available at [https://osf.io/2uzsx/]. The materials used in this study are widely available.

Code availability statement: Data analysis script notebooks are publicly available at [https://osf.io/2uzsx/].

Acknowledgments: We thank Maryam Bin Meshar, John Andrew Chwe, Benjy Barnett, Hayoung Woo, and Clodagh Cogley for assistance in materials development and data collection. This work was supported in part by research grants NIH-F31-MH114505 (R.M.S.) and NSF-BCS-1654731 (J.B.F.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions: R.M.S., E.H., and J.B.F. developed the theoretical perspective. R.M.S. developed the study concepts. All authors contributed to the study design. Testing and data collection were performed by R.M.S. R.M.S. and E.H. performed the data analysis and interpretation. R.M.S. drafted the manuscript, and all authors contributed to edits and revisions. All authors approved the final version of the manuscript for submission.

Competing interests: The authors declare no competing interests.

Supplementary Information

The entire analysis pipeline is provided and explained in Jupyter notebooks provided on our OSF page (https://osf.io/2uzsx/).

Supplementary Methods

Trait inference task instructions (all studies). These instructions were used for all tasks in which participants rated target and trait adjective stimuli in order to measure their inferences towards stimuli or their conceptual associations (included in Studies 1 – 6).

Conceptual association tasks. Specifically, participants were instructed, "In the following task, you will be presented with a series of adjective pairs. These are human personality traits. You will be asked to rate the likelihood that individuals with one of the traits possess the other trait."

Face trait task. We instructed participants, "In this task, we ask you to indicate how [TRAIT STIMULUS] a number of different people look. You will see a person's face, and are asked to judge their likely personality traits merely from their face. Importantly, go with your gut feeling. We all make snap judgments of others constantly, so feel free to report what you think about the person based on their face. Please respond quickly with your gut feeling. There are no right or wrong answers."

Familiar trait task. We instructed participants, "We are interested in personality impressions of different individuals. In this task, we ask about personality impressions of different well-known individuals, such as politicians, historical figures, and celebrities. While you may not know these individuals directly, we ask you to report how [TRAIT STIMULUS] each person is to the best of your knowledge and ability. Importantly, go with your gut feeling. We all hold snap personality impressions of others constantly, so feel free to report what you think about the person. Please respond quickly with your gut feeling. There are no right or wrong answers."

Group trait task. We instructed participants, "We are interested in the nature of stereotypes in the United States - not in studying whether people are prejudiced or not in any way, but in what common/well-known stereotypes are (these may or may not be stereotypes you yourself hold). Importantly, we are not interested in whether you endorse stereotypes or not, but instead we are interested in stereotypes that a typical American might hold. Please answer the following questions based on what you believe the stereotypes of a typical American are. In this task, we ask that you rate whether different groups of people are stereotyped as [TRAIT STIMULUS] by the typical American. Please base these ratings on what you think common stereotypes of these groups are. Please remember that stereotypes do not necessarily need to be accurate or inaccurate, negative, positive, or neutral - they just need to be widely held ideas about personality traits or behaviors in a group."

Valence task. Participants were instructed, "In this study, we would like to understand what you think about certain personality traits. You will be presented with a series of adjectives. We will have you rate each adjective on how negative or positive you believe the personality trait to be. For instance, how negative to positive is 'smart'? After reflecting on the trait word, please provide an honest response. There are no right or wrong answers." Next, participants were reminded, "In this task, you will view a series of adjectives. These are human personality traits.

Please rate the following traits based on how negative to positive they are." The matrix is provided in Supplementary Fig. 1.

Experimental study task instructions.

Conceptual association manipulation (Study 5). Upon entering the study, participants were given an overview of the study, "In this study, we are interested in how people think of others. For instance, who do you find to be kind or smart? You will complete two tasks. First, we would like to understand what you think about personality. We will have you learn about personality traits and provide your thoughts. Second, we would like to understand how you figure people out. You will be presented photos of faces and we are going to ask you a few questions about your impressions of them". After, participants read the faux scientific article (available via https://osf.io/2uzsx/). Next, participants were asked to summarize the article, "Now that you have read about human psychology and personality, we would like to hear your thoughts about the article. Please provide a summary of the article's main points, and provide a few of your thoughts about the article. In your response, provide at least a few sentences to both summarize and provide your thoughts, and remember your summary of the main points will be used as a check that you followed instructions and completed this study". Participants were debriefed to inform them of the fictitious article and its conclusions following the study.

Conceptual association through perception learning task (Study 6). For the learning phase, participants were instructed, "These instructions are very important, please read them carefully, as you will be tested for your ability to follow them. Psychologists have found that people perceive others' personality traits from their facial appearance. In one case, people are able to tell whether other people are cautious based on what they look like to various degrees. These judgments are not perfect or consistent; however, we figure them out none the less above average. In this task, we want to measure how well you are able to tell if people are cautious based upon their appearance. Remember, answers will never always be right or wrong, but just do the best you can. You will now begin the task. Please rate whether you believe the person shown is CAUTIOUS. Use the F key for 'LESS cautious' and the J key for 'MORE cautious'. You will be given feedback as to whether your answers are correct or incorrect.". Participants were debriefed to inform them of the fictitious manipulated relationship of traits following the study.

Data preparation and analysis.

Study 1.

Representational similarity analysis (RSA). All analyses were conducted with scientific and statistical libraries in Python. No subjects were removed from these data before analysis. To assess the correspondence of trait spaces across these many domains, we applied a quantitative method from systems neuroscience, RSA[1].

Each trait space may be represented as a matrix of all pair-wise similarities (e.g., correlations) between traits, or 'similarity matrix', as measured in each domain (e.g., correlation of face impressions across all trait-pairs; see Fig 2a; Supplementary Fig. 1). Each matrix may then be flattened into a vector (i.e., variable) of unique pair-wise trait similarities, by selecting values above the diagonal (thereby removing duplicate values on the opposing side of the diagonal given the similarity matrix is symmetrical, and also removing self-similar values along the diagonal). This similarity vector holds all unique information in its respective trait space similarity matrix (for an intuitive example, see[2]). Representation of each trait space matrix as a one-dimensional vector allows traditional univariate statistical methods to test the correspondence between trait space matrices. In the current research, we measure the correspondence of trait space matrices as the Spearman rank correlation between the two matrix vectors. Rank-ordering is preferable when comparing similarity matrices from different measures as it does not assume a linear relation[1]. Therefore, to conduct our analyses we computed similarity matrices per each unique trait space (conceptual, face trait impression, familiar person knowledge, group stereotype, and NEOPI trait spaces). Each similarity matrix was then converted to a vector, then values transformed into their rank position in the vector for submission to a Spearman correlation analysis to test significance.

Similarity matrices. Similarity matrices were computed for each unique trait space. Each similarity matrix was a symmetric matrix representing the pair-wise similarities between all 15 personality trait stimuli (Fig. 2a; Supplementary Fig. 1; see ‘Personality trait stimuli’ in the Methods). Excluding the conceptual trait similarity matrix, all similarity matrices were computed in the same way.

Data from each task (besides the conceptual trait task) were transformed into a format in which each trait is represented as a vector, in which its features are the level of that trait across different exemplars (per trait space, exemplars were, face: unique face stimuli, familiar person knowledge: unique familiar person stimuli, stereotype content: unique social group stimuli). For each social perceptual task, we calculated the average of each trait rating per unique stimulus to create these feature vectors per trait. Therefore, each dataset was a n (15; trait stimulus) × m (number of exemplars in that task) matrix, in which each value is the trait level of a given exemplar (e.g., ‘friendly’ vector in the face task is the ‘friendly’ rating of each face exemplar in that task). We then calculated the Pearson correlation between all trait vector pairs (Pearson correlation is used as the similarity measure to create each similarity matrix, whereas Spearman correlation is used to compare them[1]), providing the pair-wise similarity between traits as measured in each trait space matrix (a total of 105 possible unique pair-wise combinations of all trait stimuli; see Fig. 2). For the conceptual trait similarity matrix, we simply computed the mean similarity rating of each unique trait-pair, providing the full matrix.

Study 2. Study 2 applied an identical analysis as Study 1 (see above). The only difference was use of new and distinct items to represent the traits in each similarity matrix, for instance, ‘likelihood to compliment others’ and ‘likelihood to agree with others’ in place of ‘kindness’ in the matrices. Therefore, the only difference in the Study 2 analysis pipeline was the relabeling of each of the new NEOPI behavioral description items (e.g., ‘likelihood to compliment others’ and ‘likelihood to agree with others’) to their original trait terms (e.g., ‘kindness’), so that the similarity matrices could share an identical form across domains. Similarity matrices are presented in Supplementary Fig. 2).

Study 3. The face and conceptual trait space matrices were prepared in the identical strategy of that used in Study 1, however within each participant. Furthermore, we calculated the group-average conceptual trait space matrix, in order to control for consensual trait associations and target any contribution of each participants’ unique and subjective conceptual associations to their face impressions. Next, a dataset was prepared to be submitted to a multilevel mixed-effects model. In this multilevel dataset, data are cross-classified between subject and trait-pair, in which each row is a trait-pair. There are four variable columns with data per each row (specific to a subject and trait-pair): subjective face trait impression correlation, subjective conceptual association, group-average conceptual association, and valence similarity (via Study 1 data for control). Analyses were performed as a linear mixed-effects model with the lmer package in R (‘lme4’, ‘lmerTest’), applying an additional set of algorithms to assist convergence (‘brms’). All variables were *z*-normalized within participants to assist model convergence. Random slopes and intercepts were allowed for all predictor variables.

Study 4. In Study 4, we ask whether the amount to which each perceiver associates two trait concepts relates to the correlation between those trait inferences towards faces, familiar people, and groups. That is, we intended to test whether perceivers with weaker/stronger conceptual trait associations also show more weakly/strongly correlated inferences. To do so, within each perceiver, we calculated two variables: their conceptual and perceptual (face, person, or group) trait associations. To estimate their perceptual trait association, we calculated the Pearson correlation coefficient between both trait evaluations of the target stimuli within each participant (between the vectors of their inferences of all target stimuli on each of the two traits they were assigned). To estimate their conceptual trait associations, we averaged the two conceptual trait item responses. Therefore a single dataset was created including data from participants across all trait-pair combinations. Lastly, to test our hypothesis, we calculated the Spearman correlation between participant perceptual trait and conceptual trait associations (Spearman correlation used so as to not assume a strictly linear relationship between distances in the two spaces)[1]. Analyses were conducted across trait-pair terms, to assess the tendency of conceptual trait associations to relate to inference correlations, across trait-pairs in general.

Study 5. Participants who did not demonstrate a condition-consistent conceptual association were omitted for analyses (e.g., participants omitted if in the positive association direction condition they reported a neutral (4) or negative (< 4) association of their trait-pair, and vice versa for the negative association direction condition). In order to study how conceptual associations (e.g., negative vs. positive associations of ‘friendliness’ and ‘intellectualism’) of participants impact face impression correlations (e.g., lower or higher correlation of ‘friendly’ and ‘intellectual’ face impressions), we created a dataset where the participants’ subjective ratings were nested within participant, along the one trait they rated faces upon from their assigned trait pair (e.g., ‘friendly’ ratings for a participant assigned to both ‘friendly’ and ‘intellectual’). For each participant, the dependent variable was the average rating of each face (from Study 1 data) along the other trait from the participant’s trait pair they did not rate faces along (e.g., ‘intellectual’). This allowed us to estimate the relationship between each participant’s subjective perception of faces along one trait from their assigned trait-pair with the appearance of the faces along the other trait from the pair. This was done to reduce transparency and suspicion of the research goals (e.g., that we were interested in the association of the two traits they read about in the faux article, in the context of faces). Only faces were used in this study given potential limitations of the manipulation, suspicion, and social desirability in responses in the context of familiar person impressions and group stereotypes. Face impression tasks also benefit from relative unawareness from participants that they make such inferences and of how they do so[3]. Given participants rated faces along only one of the two traits to which they were assigned, to measure the face trait impression relation of the trait-pairs within each subject, the multilevel model predicted the appearance of faces along one trait (which participants did not judge; face rating data via Study 1; both studies used the same face stimuli) with participants’ subjective ratings of faces along the second trait in their assigned trait-pair. For instance, if a participant was assigned to ‘friendly’-‘intellectual’, and only rated faces on ‘intellectual’, we estimated their ‘friendly’-‘intellectual’ face impression association by predicting the average ‘friendly’ ratings of those face stimuli as measured in Study 1 with the participant’s ‘intellectual’ ratings of the face stimuli from the current study. In order to test impact of conceptual association direction condition, their assigned between-subjects condition was included as a contrast coded variable (-

1 for 'Negative', 1 for 'Positive'). Thereby, this dataset allows us to test whether participants in the positive association direction condition show higher trait inference correlations than participants in the low correlation condition. To perform this analysis, we used a multilevel mixed effects model to regress (Study 1) average ratings of the faces on (Study 5) participant subjective ratings, condition, and their interaction (analysis performed via the lmer package in R; 'lme4', 'lmerTest'). Participant ratings were group-centered. Intercepts were random but slopes were fixed.

***Study** 7*. The NEOPI trait space matrix was prepared from an open dataset (see below). Facet vectors of trait scores from many participants were Pearson correlated, measuring the similarity of actual personality traits as the correlation of these personality traits (measuring whether individuals lower/higher in one trait are lower/higher in other traits). The NEOPI trait space matrix is provided in Fig. 5.

NEOPI dataset.

All NEOPI data used to create the NEOPI trait similarity matrix (Study 7) was obtained from a publicly available dataset from prior published research[4]. Below is a summary of methods from this prior research, as well as criteria for the subset of this data utilized in the current research.

Participants. To measure the similarity structure of personality traits in the general population, we obtained a personality measurement dataset available via the Open Science Framework (OSF). In this data, a large body of participants (initial n = 334,161) completed the 300-item NEO personality test[4] (retrieved from https://osf.io/tbmh5) via a public website (http://www.personal.psu.edu/~j5j/IPIP/). In accordance with previous validity standards (publicly available by the author at http://ipip.ori.org[5]), participant responses were filtered for duplication, insufficient attentiveness, missing responses, and weak internal consistency (final n = 307,313; M_{age} = 25.2 years, SD_{age} = 10.0 years; 185,149 Female, 122,164 Male; race/ethnicity data unavailable).

Stimuli. Participants completed the 300-item NEOPI, used to measure the 30 facet subscales of the five-factor model[4]. This included a total of 300 items, with 10 items pertaining to each of the 30 total subscales of the NEOPI. An important limitation in studying the overlap of trait spaces is that semantic similarity between trait adjectives used in each task may contribute to trait inference correlation structure[6]. The strength of using the NEOPI to measure actual personality structure is its use of a wide range of self-descriptions to measure each personality trait. Rather than asking participants if they perceive themselves as each adjective (used in the social perception tasks, e.g., 'trustworthy'), participants rated themselves on several self-descriptions that correspond to the personality construct in question (e.g., they indicate the degree to which they, "Believe that others have good intentions" or "Suspect hidden motives in others"). This mitigates the possibility that a similarity in NEOPI trait space is similar to social cognitive trait spaces merely due to participants answering semantically related items similarly.

Protocol. Participants completed the 300-item NEOPI, used to measure the 30 facet subscales of the five-factor model[4]. Participants were first given instructions, "The following pages contain phrases describing people's behaviors. Please use the rating scale next to each phrase to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then click the circle that corresponds to the accuracy of the statement", followed by general protocol instructions and informed consent. Following, participants completed the 300-items in randomized order (1 - "Very inaccurate" – 5 - "Very accurate" Likert scales), in five block sets of 60 items each.

Supplementary Results

Confirmation of a common trait space. A qualitative observation of prior work[7] is that trait space models are similarly structured across social perceptual domains. This is an important assumption of the present work, as if social perceptual trait spaces reflect the conceptual trait space, they should share structure across domains. We tested this assumption quantitatively, finding a high degree of similarity across all social cognitive trait space matrices (face – familiar person trait space matrices, Spearman $\rho(103) = 0.841$, $\rho^2(103) = 0.707$, $p < 0.0001$; 95% CI = [0.774 , 0.889]; face – social group trait space matrices, Spearman $\rho(103) = 0.794$, $\rho^2(103) = 0.631$, $p < 0.0001$; 95% CI = [0.711 , 0.856]; familiar person – social group trait space matrices, Spearman $\rho(103) = 0.824$, $\rho^2(103) = 0.679$, $p < 0.0001$; 95% CI = [0.751 , 0.877]). These results confirm the assumption that there is indeed a common trait space in social perception[7]. To our knowledge, these results also provide a first quantitative assessment of the commonality between social perceptual trait spaces.

The relationship of conceptual and social perceptual trait space while controlling for valence.

Study 1. Given the apparent clustering by valence of traits in the similarity matrices (Fig. 2), we also conducted these analyses as a multiple linear regression controlling for the valence similarity matrix (based on the absolute difference of valence ratings of each trait term; $n = 69$; Supplementary Fig. 1). We found social perceptual trait space matrices were each predicted significantly by the conceptual trait space matrix over and above the valence matrix (conceptual matrix predicts: face trait space matrix, $t(102) = 7.049$, $p < .0001$, $r^2 = .328$, 95% CI = [0.185, 0.330]; familiar person trait space matrix, $t(102) = 6.553$, $p < .0001$, $r^2 = .296$, 95% CI = [0.179, 0.334]; social group trait space matrix, $t(102) = 6.910$, $p < .0001$, $r^2 = .320$, 95% CI = [0.184, 0.333]). As we would expect, we also found that valence similarity significantly predicted these social perceptual trait matrices as well (and in person knowledge and stereotypes, the effect sizes were smaller than predictions from conceptual trait space similarity; valence matrix predicts: face trait space matrix, $t(102) = 7.086$, $p < .0001$, $r^2 = .330$, 95% CI = [0.116, 0.206]; familiar person trait space matrix, $t(102) = 4.199$, $p < .0001$, $r^2 = .147$, 95% CI = [0.054, 0.151]; social group trait space matrix, $t(102) = 5.383$, $p < .0001$, $r^2 = .221$, 95% CI = [0.079, 0.172]). Indeed, valence has long been noted to be a major factor in the organization of social perceptions (e.g., even used as an alternative labeling to the 'trustworthiness' dimension in the two-factor model of face impressions[8]). Our theoretical account is agnostic to the valenced nature of particular traits. A similar or dissimilar valence among two traits surely would play a role in driving traits' conceptual similarity, but in our view it is only one contributor. By demonstrating strong effects of conceptual trait space after controlling for the contribution of valence (and equal if not stronger effects of conceptual trait space than valence space), the results cast doubt on the possibility that purely affective associations are driving the observed effects.

Study 2. In Study 2, effects of conceptual matrices on social perceptual matrices remained significant when controlling for the valence matrix, taken from Study 1, in multiple regression (conceptual matrix predicts: face trait space matrix, $t(102) = 3.196$, $p = .002$, $r^2 =$

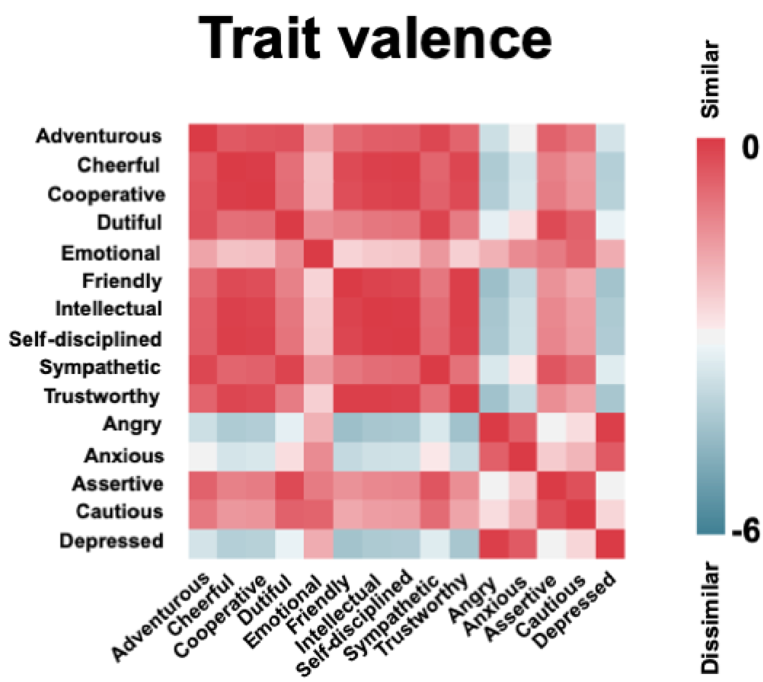
.091, 95% CI = [0.077, 0.328]; familiar person trait space matrix, $t(102) = 5.148$, $p < .0001$, $r^2 = .206$, 95% CI = [0.156, 0.352]; social group trait space matrix, $t(102) = 2.724$, $p = .008$., $r^2 = .068$, 95% CI = [0.036, 0.229]).

Study 3. In Study 3, we additionally performed the analysis controlling for the valence model collected in Study 1, finding subjective conceptual trait associations had a significant relation to face impressions over and above both group-average conceptual associations and their valenced structure ($b = 0.144$, $SE = .019$, $t(142.4) = 7.501$, $p < .0001$, 95% CI = [0.11, 0.18]).

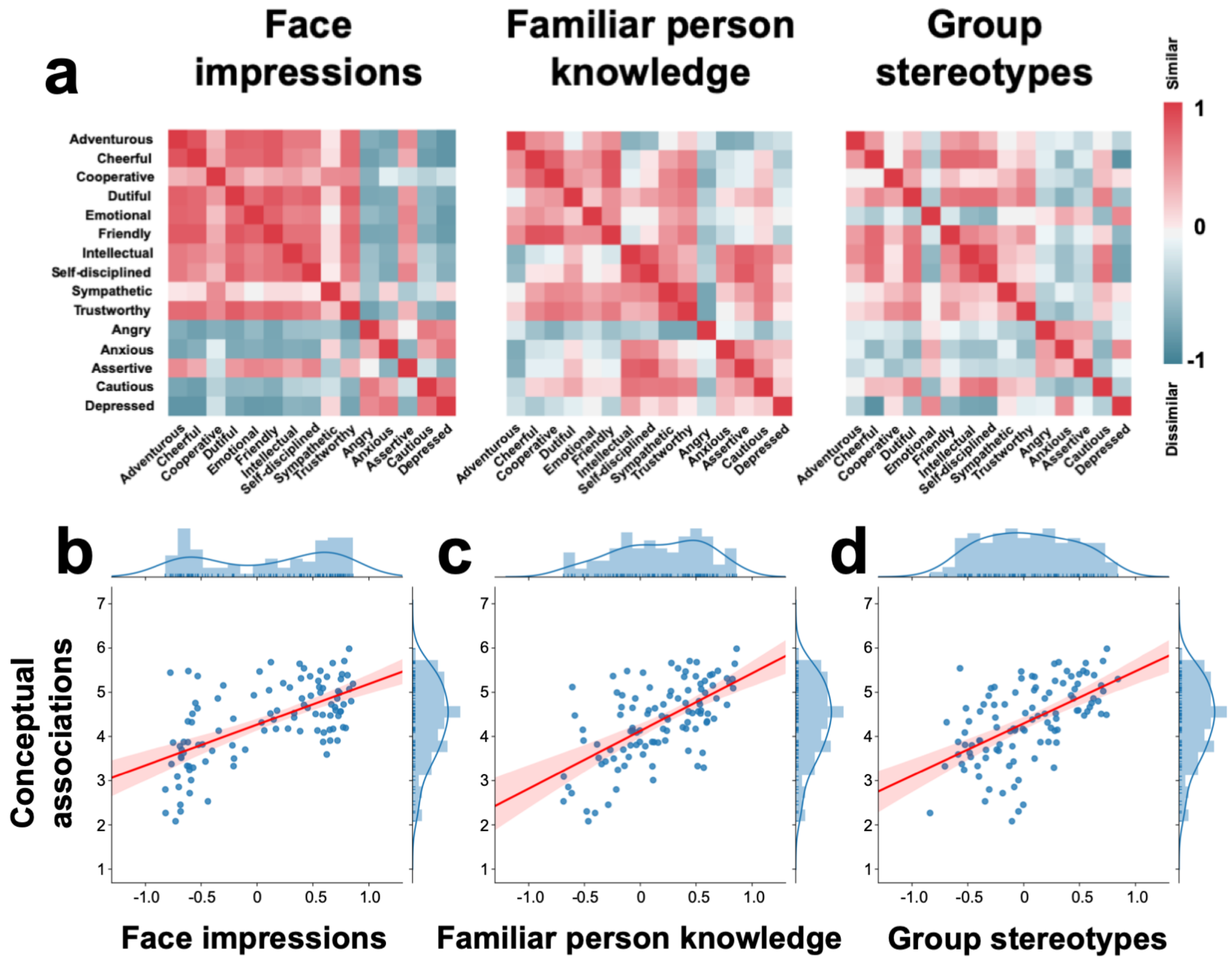
Social perceptions reflect actual personality structure. If conceptual trait space is applied to trait inferences (Studies 1-5), and perceivers' trait space mirrors actual personality structure as observed here, social perceptual trait spaces may also reflect the actual structure of personality. Indeed, we found this to be the case, as social perceptual trait space matrices from Study 1 were also strongly positively related to the NEOPI trait space matrix (NEOPI trait space matrix predicted the: face trait space matrix, Spearman $\rho(103) = 0.677$, $\rho^2(103) = 0.459$, $p < 0.0001$; 95% CI = [0.558 , 0.769]; familiar person trait space matrix, Spearman $\rho(103) = 0.644$, $\rho^2(103) = 0.415$, $p < 0.0001$; 95% CI = [0.516 , 0.744]; social group trait space matrix, Spearman $\rho(103) = 0.706$, $\rho^2(103) = 0.498$, $p < 0.0001$; 95% CI = [0.595 , 0.791]).

Additional analyses. In Studies 1, 2, and 7 we performed analyses testing the overlap between aggregate similarity matrices per each social cognitive model: conceptual, face, familiar person, and group matrices. There are two limitations of this statistical approach. In one, there is inherent dependency between elements of the matrix, as the same items and sometimes data are sometimes used to compute different elements (e.g., 'trustworthy' as input to both the 'trustworthy-angry' and 'trustworthy-anxious' cells). In another, random effects inference of any kind are not possible given matrices are aggregated across all subjects. Given entire conceptual matrices were collected per subjects in the conceptual task of Study 1, we performed additional analyses to address these limitations. These analyses were of course only possible where conceptual matrices are included in analysis. First, we conducted first-level analyses as RSA per conceptual task subject, predicting their unique conceptual matrices from the aggregate face, familiar person, and group matrices (as these were only able to be computed through aggregation across subjects in Studies 1 and 2). Second, we performed a group level analysis to test whether the similarity coefficients from first-level analyses were significant. Specifically, for each RSA reported, Spearman correlation coefficients were computed per subject, Fisher's *z* transformed, then submitted to a one-sample *t*-test against 0. Statistical test results are reported as performed on Fisher's *z* transformed Spearman correlation coefficients. For interpretation, descriptive statistics and confidence intervals of the original Spearman correlation coefficients are reported. All analyses reported were significant when tested with this method in Study 1 (conceptual – face RSA, mean $\rho = .525$, ρ SD = .202, $t(115) = 24.342$, $p < .0001$, $r^2 = .837$, mean ρ 95% CI = [0.488, 0.562]; conceptual – familiar person RSA, mean $\rho = .494$, ρ SD = .184, $t(115) = 25.760$, $p < .0001$, $r^2 = .852$, mean ρ 95% CI = [0.460, 0.528]; conceptual – group RSA, mean $\rho = .520$, ρ SD = .195, $t(115) = 25.180$, $p < .0001$, $r^2 = .846$, mean ρ 95% CI = [0.484, 0.556]), Study 2 (conceptual – face RSA, mean $\rho = .386$, ρ SD = .170, $t(115) = 22.661$, $p < .0001$, $r^2 = .817$, mean ρ 95% CI = [0.355, 0.417]; conceptual – familiar person RSA, mean $\rho = .398$, ρ SD = .143, $t(115) = 28.004$, $p < .0001$, $r^2 = .872$, mean ρ 95% CI = [0.372, 0.424]; conceptual – group RSA, mean $\rho = .378$, ρ SD = .162, $t(115) = 22.948$, $p < .0001$, $r^2 = .821$, mean ρ 95% CI = [0.348, 0.408]), and Study 7 (conceptual – NEOPI RSA, mean $\rho = .517$, ρ SD = .192, $t(115) = 26.203$, $p < .0001$, $r^2 = .857$, mean ρ 95% CI = [0.482, 0.552]).

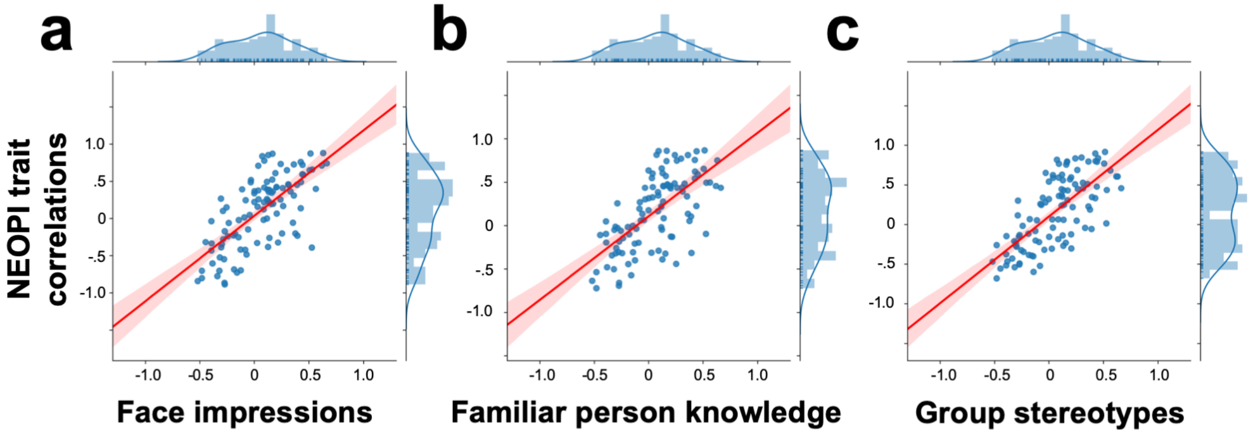
Supplementary Figures



Supplementary Figure 1. **Trait valence matrix**. In Study 1, we collected a trait space matrix of the absolute difference in valence ratings (dissimilar/blue to similar/red) of each trait adjective stimulus (*n* = 69). This was used as a control in Studies 1, 2, and 3 analyses to measure the redundancy of conceptual trait space and the valence similarity of trait terms, given the large contribution of valence to trait inferences and conceptual knowledge[8]. Control allowed analyses to measure the independent contribution of non-valence related conceptual similarities in trait-pairs to trait inferences.



Supplementary Figure 2. **Study 2 results**. First depicted are all social perceptual trait space similarity matrices from Study 2 (panel a), each made of the pairwise similarity values between each trait-pair. Each matrix is sorted by the k-means cluster solution of the conceptual trait space matrix, as to most intuitively depict their similar structure. Importantly, Study 2 used different descriptors for traits in each domain, for instance, while ‘friendly’ was used in the conceptual task, ‘likely to agree with others’ was used in face impressions. Study 2 used the same conceptual association data as Study 1 (see Figure 2, panel a). Second, we see evidence that conceptual trait space (n = 116; y-axis) substantially overlaps with social perceptual trait space across domains (x-axes; face impressions, panel b, n = 496, Spearman $\rho(103) = 0.575$, $\rho^2(103) = 0.331$, $p < 0.0001$; 95% CI = [0.431 , 0.691]; person knowledge, panel c, n = 478, Spearman $\rho(103) = 0.576$, $\rho^2(103) = 0.332$, $p < 0.0001$; 95% CI = [0.432 , 0.691]; and group stereotypes, n = 489, Spearman $\rho(103) = 0.574$, $\rho^2(103) = 0.329$, $p < 0.0001$; 95% CI = [0.430 , 0.690]). Error ribbons reflect standard error of effect estimates. While Pearson correlations are plotted for ease of interpretation, statistical analyses were of rank ordered data points. In each plot, trait space matrices (panel a) are flattened into their unique pair-wise similarity values and plotted against one another (conceptual on the y-axis, social perceptual matrices along the x-axes). Each data point is a trait-pair (e.g., ‘friendly’-‘self-disciplined’; 105 trait-pairs make up data points per panel). In each comparison, as two traits become more associated in conceptual knowledge, they become more correlated in trait inferences across domains.



Supplementary Figure 3. NEOPI trait space predicts social perceptual trait spaces. In Study 7, we find perceivers' conceptual trait associations (*y*-axes) are strikingly reflective of the actual correlation structure of personality traits (n = 307,313; *x*-axes; in face impressions, panel a, n = 484, Spearman $\rho(103) = 0.677$, $\rho^2(103) = 0.459$, $p < 0.0001$; 95% CI = [0.558 , 0.769]; person knowledge, panel b, n = 503, Spearman $\rho(103) = 0.644$, $\rho^2(103) = 0.415$, $p < 0.0001$; 95% CI = [0.516 , 0.744]; and group stereotypes, panel c, n = 488, Spearman $\rho(103) = 0.706$, $\rho^2(103) = 0.498$, $p < 0.0001$; 95% CI = [0.595 , 0.791]). Error ribbons display standard error around effect estimates, and there are 105 trait-pairs as data points per panel. While Pearson correlations are plotted for ease of interpretation, statistical analyses were of rank ordered data points. These results suggest a possibility that actual trait correlations are learned conceptually, and thereafter influence social perception. This does not necessarily entail accuracy in social perception ipso facto. This point is addressed in detail in the discussion.

Supplementary References

1 Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 4, doi:10.3389/neuro.06.004.2008 (2008).

2 Stolier, R. M., Hehman, E. & Freeman, J. B. A Dynamic Structure of Social Trait Space. *Trends in cognitive sciences* **22**, 197-200, doi:https://doi.org/10.1016/j.tics.2017.12.003 (2018).

3 Tskhay, K. O. & Rule, N. O. Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review* **17**, 72-86 (2013).

4 Johnson, J. A. Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality* **51**, 78-89 (2014).

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

6 Schneider, D. J. Implicit personality theory: A review. *Psychological bulletin* **79**, 294 (1973).

7 Fiske, S. T., Cuddy, A. J. & Glick, P. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* **11**, 77-83 (2007).

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