Research Article

Psychological Science
1–17

DOI: 10.1177/09567976211072814
www.psychologicalscience.org/PS

Personality Across World Regions Predicts Variability in the Structure of Face Impressions

DongWon Oh, Jared D. Martin, and Jonathan B. Freeman
Department of Psychology, New York University

Abstract

Research on face impressions has often focused on a fixed, universal architecture, treating regional variability as noise. Here, we demonstrated a crucial yet neglected role of cultural learning processes in forming face impressions. In Study 1, we found that variability in the structure of adult perceivers' face impressions across 42 world regions (N = 287,178) could be explained by variability in the actual personality structure of people living in those regions. In Study 2, data from 232 world regions (N = 307,136) revealed that adult perceivers use the actual personality structure learned from their local environment to form lay beliefs about personality, and these beliefs in turn support the structure of perceivers' face impressions. Together, these results suggest that people form face impressions on the basis of a conceptual understanding of personality structure that they have come to learn from their regional environment. The findings suggest a need for greater attention to the regional and cultural specificity of face impressions.

Keywords

person perception, face processing, social cognition, semantic memory, cultural psychology, open data

Received 5/21/21; Revision accepted 12/20/21

Although warned not to judge a book by its cover, people nevertheless make inferences about any number of others' personality traits on the basis of their facial appearance. These trait judgments, or *face impressions*, are made with less than 100 ms of exposure and tend to be consistent across different perceivers (for a review, see Todorov et al., 2015). Judgments of specific traits (e.g., friendliness) are highly correlated with one another; thus, the structure of face impressions can be summarized by only a few dimensions, such as trustworthiness and dominance (Lin et al., 2021; Oosterhof & Todorov, 2008; Sutherland et al., 2013). These core dimensions have often been interpreted through the lens of universal, evolutionarily adaptive processes, such as tracking other people's intentions (e.g., trustworthiness) and the ability to enact those intentions (e.g., dominance; e.g., Fiske et al., 2007; Oosterhof & Todorov, 2008). Indeed, recent studies have found that the structure of face impressions is largely consistent across world regions (Jones et al., 2021; Lin et al., 2021), lending some support for a universal structure, although between-region variability has also been observed.

Regional variability has been documented in various domains of face perception, including face impressions (Birkás et al., 2014; Zhang et al., 2019), emotion perception (Elfenbein et al., 2002; Jack et al., 2012; Soto & Levenson, 2009), and more basic face-perception processes (Caldara, 2017). However, to our knowledge, regional variability in the structure of face impressions has yet to be systematically demonstrated. Increasingly, research has documented meaningful differences across individual perceivers in their face impressions (Holzleitner & Perrett, 2017; Hönekopp, 2006; Martinez et al., 2020; Oh et al., 2020; Sutherland et al., 2020; Xie et al., 2018) and across target social categories (Collova

Corresponding Author:
DongWon Oh, New York University, Department of Psychology
Email: dongwonoh.phd@gmail.com

et al., 2019; Hehman et al., 2017; Oh et al., 2020; Sutherland et al., 2015). These findings suggest that although the notion of a universal structure of face impressions may successfully explain average trends, key variability may have gone relatively ignored. One source of this variability is the idiosyncratic difference in perceivers' conceptual beliefs about traits and their covariation (e.g., to what extent being "aggressive" relates to being "intelligent"; Stolier et al., 2018). Notably, these conceptual trait relations could be acquired through statistical learning processes as perceivers observe their social environment, including how other people's personality traits covary (Stolier et al., 2020)—a premise consistent with classic research on how we implicitly infer others' personalities (Schneider, 1973) and more recent research on the role of environmental factors (Sutherland et al., 2020) and statistical learning processes (Dotsch et al., 2016) in face impressions.

Much like conceptual relations across traits in people's minds, people's actual personality traits tend to be correlated along a small set of dimensions (e.g., the Big Five; Costa & McCrae, 1992). Because actual personality traits are highly correlated, a simple strategy for perceivers to optimize trait inference would be to learn this correlation structure and make predictions accordingly. If perceivers learn the actual structure of personality, traits that are more similar in actual human personality would become conceptually believed to be more similar. If true, this possibility suggests that the conceptual structure of personality traits (and, in turn, the structure of trait judgments of faces) would approximate the structure of actual personality traits perceivers observe in their social environments.

One important social environment may be the world region in which perceivers reside. Although the structure of personality is theorized to be universal (e.g., McCrae & Costa, 1997), reliable regional and cultural differences are often observed (McCrae, 2001, 2002). In particular, the dimensions of extraversion, agreeableness, and openness to experience have been found to vary across world regions (McCrae & Terracciano, 2005; Rolland, 2002). Thus, despite a degree of universality in the structure of personality, meaningful variability in human personality structure across world regions may shape the structure of perceivers' conceptual understanding of personality, which in turn may drive regional differences in the structure of how personality is judged from faces.

Here, we hypothesized that the average personality in perceivers' world regions would explain the conceptual understanding of personality in perceivers' minds and, in turn, explain how they infer personality in others' faces. For example, if a perceiver grows up in a world region where aggressive individuals tend to be intelligent, then the perceiver will tend to believe that aggressiveness and intelligence are conceptually related. As a result, the same perceiver will use similar facial features to judge whether targets are aggressive or intelligent and thus show positive correlation in their face judgments of these two traits. On the other hand, a perceiver who grows up in a region where aggressiveness and intelligence have little relationship would not develop this conceptual association and, in turn, would not show such a correlation in face judgments. Because this learning process occurs for all pairs of personality traits, the structure of personality in one's world region would become the structure of one's conceptual beliefs about personality, which in turn would drive their face-based personality judgments. Indeed, environmental factors and statistical learning processes play a key role in face-based personality judgments (Dotsch et al., 2016; FeldmanHall et al., 2018; Stolier et al., 2020; Sutherland et al., 2020; Verosky & Todorov, 2010).

We tested our overall hypothesis across two studies. In Study 1, using personality data from individuals across 42 world regions together with an independent data set of face-based trait judgments from the same regions, we tested whether the regional structure of personality traits predicts the regional structure of face-based trait judgments. In Study 2, we examined the intermediary role that perceivers' learned conceptual understanding about personality traits plays. Together, the results of these two studies suggest that the structure

Statement of Relevance

Prominent models of how people form impressions of personality traits from faces, such as trustworthiness, focus on universal mappings between specific facial features and traits, ignoring the role of cultural learning. In two studies, we found that the actual personalities of people living in a world region were related to how individuals in that region judge others' traits from faces. For example, in a region where people are more likely to be simultaneously aggressive and intelligent, people in that region are more likely to judge a person with an aggressive-appearing face as more intelligent and vice versa. We additionally provide evidence suggesting that people use this actual personality structure learned from their local environment to form a conceptual understanding about human personality, which then drives how they form impressions of faces. The findings point to a crucial role of cultural learning in first impressions of faces.

of people's actual personalities in a given world region shapes the conceptual understanding of personality in that region, which in turn affects how trait impressions of faces are formed in the region.

Study 1

Using international data sets of self-reported personality inventories and face-based trait judgments across 42 world regions, we tested whether people's self-reported personalities were related to how individuals in those regions judge others' personalities from faces.

Method

Participants. For the personality data, we used self-reported personality ratings from multiple world regions (Johnson, 2014; data are available at https://osf.io/wxvth). Online participants living in 232 different world regions (N = 307,313) participated in a personality survey. For face-impressions data, 13,671 total participants living in 43 world regions participated in a laboratory setting (Jones et al., 2021; data available at https://osf.io/f7v3n). The research protocols used for data collection by each research group were approved by their local ethics committee or institutional review board. We used only the subset of face-impressions data that corresponded to the same regions as those of the personality data. This resulted in final samples of self-reported personality data from 287,178 participants in 42 regions (age: M = 25.22 years, SD = 10.05; 39.24% male, 60.76% female) and face-judgment data from 13,671 participants in those same 42 regions (age: M = 22.63 years, SD = 7.00; 29.12% male, 69.60% female, 0.73% declined to report gender or reported another gender). The final 42 world regions were geographically and culturally diverse (see Fig. S1a in the Supplemental Material available online for a complete list and the locations of the regions). All participant samples were convenience samples. We used all available participants' data, and we did not predetermine the sample size.

NEO Personality Inventory (NEO-PI) data. Participants answered the 300 items of the International Personality Item Pool (IPIP) representation of the Revised NEO-PI (NEO-PI-R), a well-known personality inventory administered online (Johnson, 2014). Each of these 300 items describes a person's affective, behavioral, and/or cognitive tendency, with each item contributing to one of the six facets that compose each of the Big Five factors (*agreeableness*: morality, altruism, cooperation, modesty, sympathy, trust; *conscientiousness*: self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, cautiousness; *extraversion*: friendliness, gregariousness, assertiveness, activity level, excitement-seeking, cheerfulness; *neuroticism*: anxiety, anger, depression, self-consciousness, immoderation, vulnerability; *openness to experience*: imagination, artistic interests, emotionality, adventurousness, intellect, liberalism). For example, "Worry about things" measures the anxiety facet of the neuroticism factor, "Often feel blue" measures the depression facet of the neuroticism factor, "Prefer variety to routine" measures the adventurousness facet of the openness-to-experience factor, and "Like to get lost in thought" measures the imagination facet of the openness-to-experience factor. Each participant used a 5-point scale to rate how accurately each item described themselves (1 = *very inaccurate*, 5 = *very accurate*). Details of the 300-item international NEO-PI procedure are described by Johnson (2014). Mean NEO-PI scales averaged across participants within specific world regions have been found to convey meaningful region-specific information (Allik et al., 2017).

Full details on data-exclusion procedures are provided by Johnson (2005). Participants were instructed not to skip multiple responses, not to consecutively use the same response multiple times, and not to respond randomly. Randomness of responses was determined by within-participant reliability (correlation of nonoverlapping subsets of a participant's responses that corresponded to one other in meaning). Participants were excluded if they did not follow the instructions. These exclusions aimed to remove participants who did not understand the questions or were not paying attention. The IPIP-NEO was administered in English across all regions, and all participants indicated understanding test instructions and the purpose of the test. Thus, participants with poor English comprehension were excluded.

Face-impressions data. For face-based trait judgments, participants judged 120 faces on 13 personality traits. The faces were standardized photos from the well-validated Chicago Face Database (Ma et al., 2015), including 30 Asian, 30 Black, 30 Hispanic, and 30 White faces (half male and half female within each race). Each participant was asked to rate the 120 target faces, one at a time, on one of the 13 personality traits: aggressiveness, attractiveness, caringness, confidence, dominance, emotional stability, unhappiness, intelligence, meanness, responsibility, sociability, trustworthiness, weirdness (taken from Oosterhof & Todorov, 2008). On each trial, a 9-point scale with a prompt was presented below the face (e.g., "How [aggressive] is this person?); responses ranged from 1 (*not at all [aggressive]*) to 9 (*very [aggressive]*). In each region, 25 or more raters were recruited to rate faces on each of the 13 traits for a sufficient level of interrater reliability. The task in each data-collecting laboratory used the official language of their region (e.g., Farsi in Iran) or the most widely used language in the region (e.g., English in the United States) to allow all raters to complete the task in their native language. In each region, the data-collecting teams translated the trait terms and the task instructions from an initial English version with the help of English-language dictionary definitions denoting the intended meaning of each of the trait words used. This approach had been used in prior studies that tested for cultural differences in face processing (Han et al., 2018). Full details of the face-based trait-rating procedure can be found in Jones et al. (2021).

Language covariates. Previous studies have consistently found that administering a NEO-PI personality inventory in two different languages (e.g., English and a non-English native language) to the same group of multilingual individuals produces highly similar individual NEO-PI scores and regional NEO-PI structure (e.g., Church & Katigbak, 2002; Gülgöz, 2002; McCrae, 2001; McCrae et al., 1998; Piedmont & Chae, 1997; Piedmont et al., 2002; Simakhodskaya, 2000). For instance, when participants across multiple world regions completed the NEO-PI-R in two different languages about a week apart, a high test-retest reliability between the languages was observed (median r = .86; McCrae, 2001). These findings suggest that, although English language was used for participants across all regions in the NEO-PI personality data from Johnson (2014), it is unlikely to have introduced any meaningful biases in results (McCrae, 2001). The concern is further alleviated by the fact that, as described earlier, participants were excluded if their responses showed evidence of inconsistency or a lack of comprehension.

Nevertheless, it remains a possibility that regional differences in how well nonnative English speakers comprehended the English-language NEO-PI inventory could have confounded the results. For example, if most participants in a region did not understand specific questions because of their limited facility with English, or if they misunderstood the questions, this could result in biases in responses. It is also possible that similarity between any two regions' primary languages could confound effects of NEO-PI similarity on face-judgment similarity. For instance, rather than similarity in the face-judgment structure of Germany and France being attributable to corresponding similarity in the two regions' personality structure, it is possible that these effects may be better attributed to the similarity of the German and French languages (e.g., suggesting that similarity in the face-judgment data reflects similar linguistic processes in interpreting the trait words and task rather than any genuine effect of the two regions' NEO-PI structure). To eliminate these possibilities, we repeated all regression analyses so as to include several language-related covariates.

In the case of analyses conducted at the level of traits (trait-level analysis; see the Analytic Approach section), we included regions' primary language and level of English proficiency (using two complementary measures) in regression models. In the case of analyses conducted at the level of regions (region-level analysis), we included a measure of language dissimilarity between the primary language spoken in each pair of regions, as well as dissimilarity in the level of English proficiency in each pair of regions (using the two measures).

Pairwise language-distance measures were derived from the Automated Similarity Judgment Program (ASJP; Holman et al., 2008b). The ASJP database contains a set of common words across more than 7,000 languages throughout the globe (Søren et al., 2020). Language-relatedness data, such as the ASJP language distance, have been found to be capable of reconstructing the evolution of human language and culture (Atkinson et al., 2008; Pagel et al., 2007; Pompei et al., 2011) and are associated with the geography of regions in which languages are spoken (e.g., distance from water; Bentz et al., 2018). Using ASJP data for all available words with respect to the primary language spoken in each region, we calculated the Levenshtein distance for each pair of regions. Levenshtein distance is the standard method for calculating dissimilarity of languages (Holman et al., 2008a) and is based on the distance between pairs of words that have identical meanings. Specifically, it quantifies the difference between two strings, as defined by the minimum number of edited letters (i.e., insertion/deletion/substitution) needed to transform one string to the other (e.g., *blood* and *sangre*). As is common practice, after calculating the Levenshtein distance for all available words in the ASJP database on the basis of the primary language for each pair of regions, we corrected them for word length and generated a normalized Levenshtein distance measure (as longer words would lead to an unwarrantedly larger dissimilarity value; Holman et al., 2008a). All 42 regions had their primary languages in the ASJP database (24 languages in total), which allowed us to calculate dissimilarity values between all 276 language pairs.

We also included two complementary measures of regions' English proficiency in regression models: the regional average Test of English as a Foreign Language (TOEFL) score (Educational Testing Service, 2021) and the regional average English Proficiency Index (EPI) score (Education First, 2020). For region-level analyses, we included the dissimilarity (i.e., difference score) in the TOEFL score and in the EPI score between each pair of regions. Both scores are based on large numbers of test takers (over 1 million each) across the globe, allowing us to approximate each region's average level of facility with English. The most recent TOEFL and EPI reports provided the measures on 165 and 100 regions, respectively. Among the 42 target regions considered in Study 1, all 42 regions had TOEFL scores available (100% of all regions), and 36 regions had EPIs available (85.71%).

Ethnic-diversity covariate. We also considered the ethnic diversity of the population in each region. Exposure to varying levels of ethnic diversity in each region could, in theory, affect perceivers' face judgments (e.g., Birkás et al., 2014; Hills & Pake, 2013; Xie et al., 2018; Zhang et al., 2019), particularly faces that vary in ethnicity, as in the data from Jones et al. (2021). We used the Herfindahl-Hirschman Index (HHI; Hirschman, 1945), which is a measure of homogeneity in a given group, derived from a regional ethnic-fractionalization index (a probability of two randomly picked individuals belonging to two different ethnic groups; Alesina et al., 2003). The ethnic-fractionalization HHI is correlated with regional differences in face-related variables, such as emotional expressivity (e.g., Rychlowska et al., 2015). The HHI was available for all 42 regions.

Analytic approach. To test whether the structure of personality traits of people in different regions predicts the structure of personality traits judged from faces in those regions, we took a representational similarity analysis (RSA) approach that tested the correspondence between regions' self-reported NEO-PI personality trait space and face-judgment trait space. The NEO-PI space was represented by an NEO-PI trait-dissimilarity matrix comprising all pairwise dissimilarities between self-reported personality traits in each world region (i.e., how similarly or dissimilarly people rate themselves in terms of their personality). The face-judgment space was represented by a face-based trait-dissimilarity matrix comprising all pairwise dissimilarities between evaluated trait dimensions of the set of 120 faces for each region (i.e., how similarly or dissimilarly people rate others' faces).

To map the NEO-PI and face-judgment space via RSA, we analyzed only those traits common to both spaces. Two out of the 13 traits used in the face-judgment task were excluded. Attractiveness was excluded because, unlike the other traits, it refers to physical characteristics rather than inferred personality, and weirdness was excluded because it is not captured well by a high or low score on any single NEO-PI personality trait. For each region, we averaged face ratings across all participants on each of the 11 remaining traits. We then mapped the NEO-PI self-reported personality items (300 items) to the traits used for face judgments (11 traits).

Mappings were created using two converging approaches. In the first approach, the second author and four research assistants served as five coders, who for each NEO-PI item marked which (if any) of the 11 face traits best described the item and in which direction it was related to the item (positive or negative). "None of these options" was included as the last option to avoid any imprecise mapping. For each NEO-PI item, when a majority (three or more) agreed that an item corresponded to one of the 11 traits (and in the same coding direction), we considered the NEO-PI face coding of that particular NEO-PI item as conclusive. A total of 124 final NEO-PI items reached such agreement. For instance, "Like to solve complex problems" was coded as "intelligent" in the positive direction and "Avoid difficult reading material" was coded as "intelligent" in the negative direction.

To seek converging evidence, in the second approach, we recruited independent raters from Amazon Mechanical Turk living in the United States ($N = 49$). Raters were asked to indicate which (if any) of the 11 face traits best described each of the 300 NEO-PI items (including a "None of these options" response). Although the coders of the first approach additionally rated positive versus negative coding direction, we excluded these additional ratings for the independent raters. Including coding direction would have required an additional 300 responses per rater, which was infeasible given the time constraints of an online Mechanical Turk study. If a majority (> 50%) of raters agreed that an item corresponded to one of the 11 traits, we considered that NEO-PI face mapping to be conclusive. Coding direction (positive or negative) for conclusive items was taken from the in-lab coder data and was self-evident (e.g., "Lose my temper" clearly corresponds to "aggressive" rather than "not aggressive").

We included data only from independent raters who followed instructions and passed all attention-check trials (18 trials randomly interspersed across 318 total trials; e.g., "Select the option comprised of four words," "Select the second option from the bottom"). Because our aim was to extract reliable mappings, we adopted a high criterion of 100% accuracy in attention checks for data inclusion. This procedure left us with 24 raters (age: $M = 40.88$ years, $SD = 11.71$; 58.33% male, 41.67% female; 8.33% Black, 12.50% Hispanic, 4.17% Native American, 75.00% White). Among these raters, a majority (i.e., ≥ 12) reached consensus that 103 NEO-PI items reliably corresponded to one of the 11 traits.

The coders and independent raters showed substantial agreement in their mappings. Among 218 of 300 items (72.76%), the two groups agreed that an item corresponded to the same trait (or did not correspond to any trait). There was only one item for which the two groups mapped differently on a trait: "Get angry easily" was judged by the coders as best corresponding to "aggressive" but by the independent raters as "emotionally unstable." See Table S1 in the Supplemental Material for the complete NEO-PI face-coding scheme.

We averaged, for each region, all individual NEO-PI responses across participants. For each of the 42 regions, we prepared dissimilarity matrices that represented a NEO-PI space and a face-judgment space. In the 42 region-specific NEO-PI dissimilarity matrices, we calculated the euclidean distance for every pair of traits using average trait scores of all respondents on all 11 personality traits. As a result, each cell corresponded to the extent to which on average a trait pair co-occurred in individuals' self-reported personality in that particular region (e.g., co-occurrence of aggressiveness and intelligence; Fig. 1a). In the 42 region-specific face-judgment dissimilarity matrices, we calculated the euclidean distance between the 120 face-trait ratings (averaged across all participants in the region) for every pair of traits. As a result, each cell corresponded to the extent to which people's face judgments of the two traits (e.g., judgments of aggressiveness and intelligence) tended to covary (Fig. 1a). Both dissimilarity matrices, for each region, were 11 × 11 matrices, in which cells represented all pairs of 11 total traits (Figs. 1a and 1c). A larger value in any matrix indicated a stronger dissimilarity (i.e., greater euclidean distance). As is customary in RSA, we rank-ordered similarity values prior to entering them into regression models (or in correlation analyses, we used Spearman rank-ordered correlations) so as to not assume linear relationships between variables (Kriegeskorte et al., 2008). The 55 unique trait pairs—the unique values under the diagonal in the NEO-PI dissimilarity matrices and face-judgment dissimilarity matrices—were vectorized and entered in regression analyses testing the relationship between face-judgment dissimilarity values and NEO-PI personality dissimilarity values. To appropriately account for the multilevel nature of the data (55 trait pairs nested in each of 42 regions), we conducted multilevel regressions using generalized estimating equations (GEE; Liang & Zeger, 1986). For all GEE models, we report unstandardized regression coefficients and use Wald Z as a measure of effect size. For ease of interpretation, prior to analyses, all variables were rescaled to vary between 0 and 1 so that 0 corresponded to the smallest distance (maximum similarity between regions) and 1 corresponded to the largest distance (minimum similarity between regions).

We also conducted complementary RSA at the level of regions. We again mapped across a NEO-PI personality trait space and a face-based trait space, but this time with 42 × 42 dissimilarity matrices in which each cell represented the dissimilarity between any given pair of regions (Fig. 2). NEO-PI and face-judgment dissimilarity values between pairs of regions were calculated as the euclidean distance between the two regions' aggregated values for the 11 personality traits (i.e., the 13 traits of the face-judgment data after excluding attractiveness and weirdness and using the coded mappings of those 11 traits to the NEO-PI items described above). However, this region-level RSA permitted greater flexibility in testing multiple indices of NEO-PI dissimilarity and face-judgment dissimilarity, because correspondence did not need to be evaluated at the level of individual traits (only at the level of regions). Thus, to evaluate the robustness of the effects, we also calculated the NEO-PI 42 × 42 dissimilarity matrix using dissimilarity between pairs of regions in (a) the five NEO-PI factors, (b) the 30 NEO-PI facets, and (c) the full 300 NEO-PI items. The face-judgment 42 × 42 dissimilarity matrix was also calculated using dissimilarity between pairs of regions in the full 13 traits (reincluding attractiveness and weirdness). Notably, the additional analyses using all 300 available NEO-PI items (a–c) and the additional analysis using all 13 face traits ensured that the effects of interest did not depend on any specific personality–face mappings applied to our data (i.e., by the in-lab coders or independent raters).

To test the relationship between the 42 × 42 NEO-PI dissimilarity matrix and the 42 × 42 face-judgment dissimilarity matrix in each of these cases, we vectorized the 861 unique values under the diagonal of the dissimilarity matrices and assessed their Spearman correlation. Unlike with the trait-level analyses, the data were not multilevel and thus did not require GEE regression; however, for direct statistical comparison, we complemented Spearman correlations with GEE regressions for the region-level RSA.

All regression models were repeated after including the additional language-use and ethnic-diversity covariates described earlier.

Results

Trait-level analyses. We conducted a series of complementary multilevel regression analyses to provide evidence that people's unique personality structure in different world regions is reflected in how people form trait judgments of faces in those regions. First, we regressed regions' NEO-PI dissimilarity values for the 55 trait pairs onto their face-judgment dissimilarity values using GEE regression (trait pairs nested within regions). There was a strong positive relationship, regardless of whether we used mappings derived from coders ($b = 0.12$, $SE = 0.02$, 95% confidence interval [CI] = [0.08, 0.16], $Z = 5.59$, $p < .001$) or independent raters ($b = 0.06$, $SE = 0.03$, 95% CI = [0.01, 0.11], $Z = 2.21$, $p = .027$), showing that the structure of people's personalities in a region was reflected in the structure of that region's face-based trait judgments (Fig. 1). For example, if aggressiveness and intelligence tend to co-occur more in the personalities of people in a given region, then people

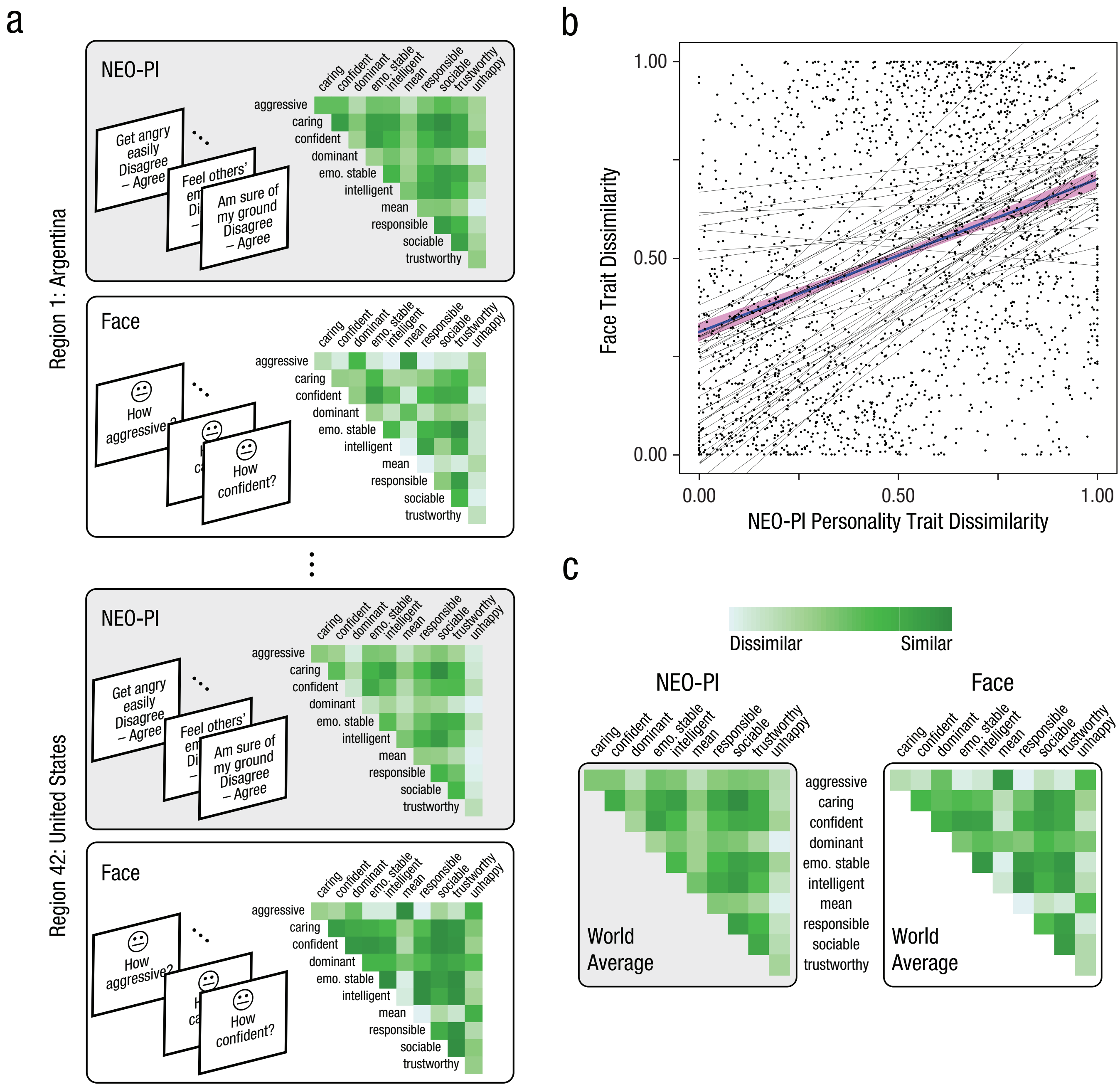

Fig. 1. Analytic approach and results of Study 1's trait-level representational similarity analysis. Two dissimilarity matrices were created for each of 42 world regions; matrices from two of these regions are shown in (a). A first set of participants (N = 287,178) answered questions from the NEO Personality Inventory (NEO-PI) about their personality (data from Johnson, 2014). An independent group of participants (N = 13,671) judged a set of 120 faces on personality traits (data from Jones et al., 2021). The euclidean distance between each pair of traits served as a measure of dissimilarity. Unique dissimilarity values in the matrices were vectorized and entered into multilevel models predicting the structure of face-based trait impressions from the structure of NEO-PI personality in the same regions. The scatterplot (b) shows the relationship between dissimilarity values of NEO-PI trait pairs and face-judgment trait pairs. Dots indicate individual trait pairs (e.g., aggressive–intelligent), thinner lines indicate slopes for individual regions, the thicker line indicates average linear fit across regions, and the shaded area represents the 95% confidence interval for the average linear fit (shown for illustrative purposes only; actual analyses were run using generalized-estimating-equations multilevel regression). For illustrative purposes, we rescaled both x and y coordinates to [0, 1] within each region; actual analyses were run using rank-ordered values and z-normalized values. The mean distances in personality traits and mean distances in face-based trait ratings, averaged across regions, are shown in (c). In (a) and (c), only the upper triangles of the matrices are displayed to avoid redundancy. Emo. stable = emotionally stable.

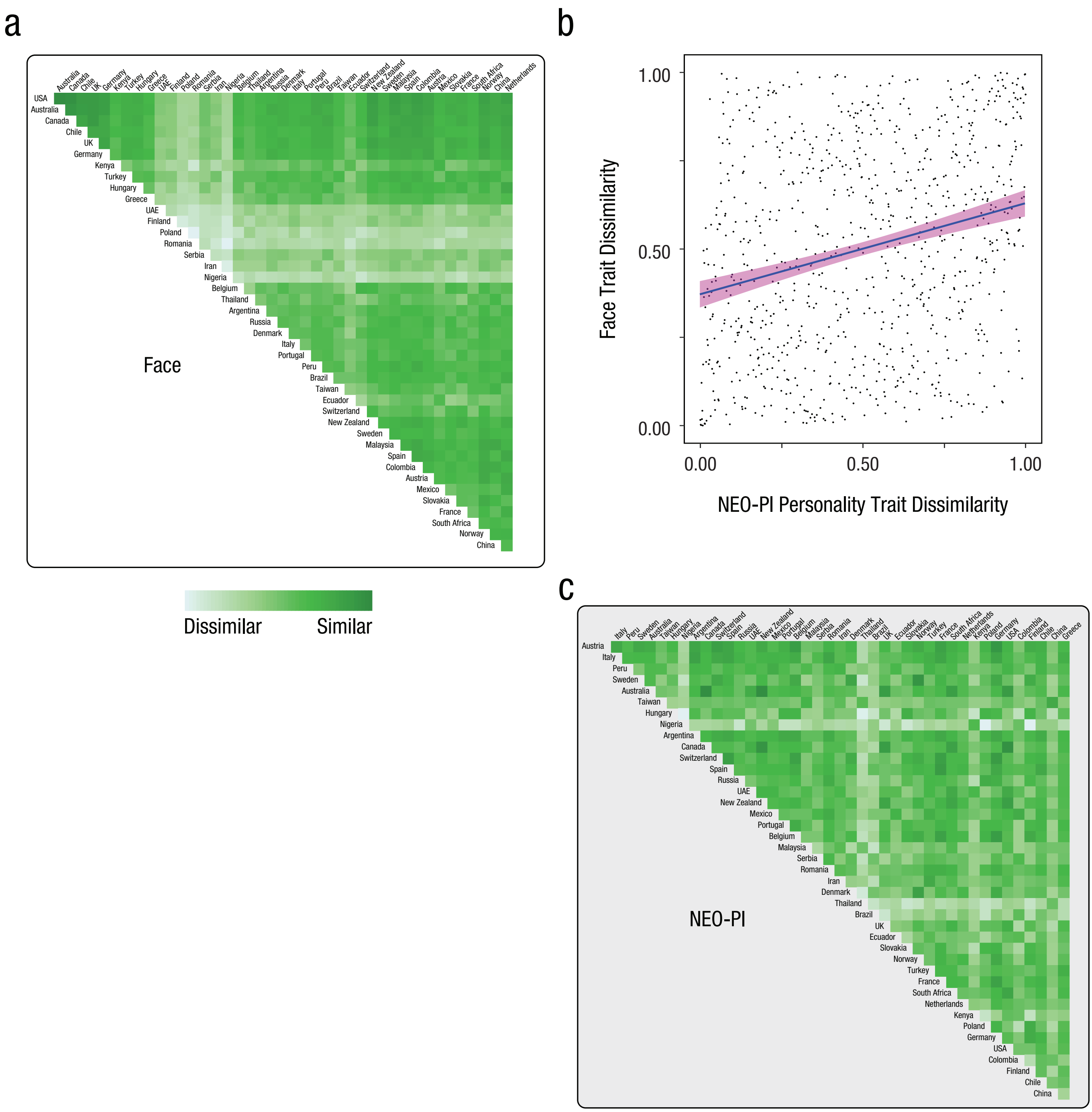

Fig. 2. Results of Study 1's region-level representational similarity analysis. Between-region similarity and dissimilarity between trait judgments of a set of 120 faces (N = 13,671) is shown in (a). Between-region similarity and dissimilarity between personality traits, assessed on the NEO Personality Inventory (NEO-PI; N = 287,178), is shown in (c). The scatterplot (b) shows the relation between face-trait judgments and NEO-PI scores, separately for each region pair (e.g., Australia–Iran). The line indicates the average linear fit across all region pairs, and the shaded area represents the 95% confidence interval for the average linear fit. For illustrative purposes, we rescaled both x and y coordinates to [0, 1] within each region; actual analyses were run using rank-ordered values. In (a) and (c), only the upper triangles of the matrices are displayed to avoid redundancy. UAE = United Arab Emirates.

also tend to evaluate aggressiveness and intelligence more similarly in others' faces in that region.

To more directly assess unique and idiosyncratic differences in NEO-PI and face-judgment structure across region, we conducted an additional multilevel regression analysis that clustered the data by trait pair instead of region. This analysis thereby was intended to show that, within a given trait pair (e.g., aggressiveness and intelligence), regions with higher NEO-PI dissimilarity values tend to also be the regions with higher face-judgment dissimilarity values for that specific trait pair. This analysis therefore serves as a more stringent test of unique interregional differences in NEO-PI structure that may be reflected in regions' face-judgment structure. NEO-PI and face-judgment dissimilarity values were *z*-normalized within each region, thereby removing any differences in magnitude or scale in these variables (i.e., the possibility that some regions have higher or lower dissimilarity values overall, or more or less dispersion, across all 55 trait pairs). Using GEE regression (regions nested within trait pairs), we regressed trait pairs' NEO-PI dissimilarity values for the 42 regions onto their face-judgment dissimilarity values, which revealed a strong positive relationship regardless of whether mappings were derived from coders (b = 0.42, SE = 0.02, 95% CI = [0.38, 0.45], Z = 23.75, p < .001) or independent raters (b = 0.43, SE = 0.02, 95% CI = [0.39, 0.48], Z = 18.88, p < .001; see Fig. S2 in the Supplemental Material).

We reconducted our analyses, this time including four covariates accounting for regions' language use and ethnic diversity: primary language, EPI, TOEFL, and ethnic-fractionalization HHI. Inclusion of these covariates did not meaningfully change the relationship between personality and face impressions. Specifically, the effects of NEO-PI personality structure on face-impressions structure remained strongly significant, regardless of whether we used mappings derived by coders (clustered by region: b = 0.22, SE = 0.02, 95% CI = [0.18, 0.27], Z = 9.17, p < .001; clustered by trait pair: b = 0.42, SE = 0.02, 95% CI = [0.38, 0.46], Z = 21.52, p < .001) or derived by independent raters (clustered by region: b = 0.18, SE = 0.03, 95% CI = [0.12, 0.24], Z = 5.66, p < .001; clustered by trait pair: b = 0.42, SE = 0.03, 95% CI = [0.37, 0.47], Z = 16.71, p < .001). See Table S2 in the Supplemental Material for full statistics.

These results show, for example, that if aggressiveness and intelligence tend to co-occur in people's personalities more in Australia than in Iran, then people in Australia also tend to evaluate the aggressiveness and intelligence of faces more similarly than do people in Iran. These complementary analyses therefore provide strong evidence that unique differences in human personality across world regions are reflected in corresponding differences in how people in those regions judge personality traits in others' faces.

Region-level analyses. As a corroborating analysis, we conducted region-level RSA, mapping NEO-PI personality trait space and a face-based trait space by region rather than individual traits using 42 × 42 dissimilarity matrices, with each cell representing the dissimilarity between any given pair of regions on the basis of the 11 personality traits (Fig. 2). Vectorizing the 861 unique values in the dissimilarity matrices, we observed a strong positive relationship between the NEO-PI dissimilarity matrix and the face-judgment dissimilarity matrix, regardless of whether we used mappings derived from coders (b = 0.26, SE = 0.03, 95% CI = [0.20, 0.32], Z = 8.04, p < .001; Spearman's ρ = .26, 95% CI = [.19, .32], p < .001) or independent raters (b = 0.24, SE = 0.03, 95% CI = [0.18, 0.30], Z = 7.55, p < .001; Spearman's ρ = .24, 95% CI = [.18, .30], p < .001). When we included the four covariates capturing dissimilarity in linguistic and ethnic diversity between region pairs—ASJP language distance, EPI difference, TOEFL difference, and HHI ethnic-diversity difference—the results did not meaningfully change. Specifically, the relationship between NEO-PI structure and face-impressions structure remained strong and significant when the four covariates were included, whether we used mappings derived by coders (b = 0.19, SE = 0.04, 95% CI = [0.12, 0.27], Z = 4.92, p < .001) or independent raters (b = 0.21, SE = 0.04, 95% CI = [0.13, 0.28], Z = 5.32, p < .001). See Table S3 in the Supplemental Material for full statistics.

At this region level, RSA does not require trait-level correspondence across NEO-PI and face-judgment space; this permitted greater flexibility to demonstrate the robustness of this relationship in a manner that did not require any personality–face mappings whatsoever. The strong positive relationship persisted regardless of whether the NEO-PI dissimilarity matrix was calculated using pairwise regional dissimilarity on the basis of the five NEO-PI factors (b = 0.13, SE = 0.03, 95% CI = [0.06, 0.19], Z = 3.76, p < .001; Spearman's ρ = .20, 95% CI = [.14, .27], p < .001); the 30 NEO-PI facets (b = 0.18, SE = 0.03, 95% CI = [0.11, 0.24], Z = 5.40, p < .001; Spearman's ρ = .18, 95% CI = [.11, .24], p < .001); or the full 300 NEO-PI items (b = 0.24, SE = 0.03, 95% CI = [0.17, 0.30], Z = 7.33, p < .001; Spearman's ρ = .23, 95% CI = [.17, .30], p < .001), or when the face-judgment dissimilarity matrix was calculated on the basis of the full 13 traits (after reincluding attractiveness and weirdness) using mappings by coders (b = 0.23, SE = 0.03, 95% CI = [0.17, 0.29], Z = 7.16, p < .001; Spearman's ρ = .23, 95% CI = [.17, .29], p < .001) or independent raters (b = 0.23, SE = 0.03, 95% CI = [0.17, 0.29], Z = 7.10, p < .001; Spearman's ρ = .23, 95% CI = [.16, .29], p < .001). Thus, the region-level RSA demonstrated a highly robust relationship between NEO-PI structure and face-impressions structure across regions. The extent to which any two regions' (e.g., Australia's and Iran's) personality structure was more similar predicted a corresponding similarity in perceivers' face-trait structure in those two regions.

Discussion

Across two types of RSA conducted at multiple levels of analysis (trait and region level), the results show that unique variability in the structure of human personality across world regions is reflected in the structure of how people in those regions form trait impressions of others' faces. Moreover, these effects held even when models accounted for regional variability in language use and ethnic diversity.

Study 2

We have hypothesized that perceivers' conceptual understanding of personality traits may explain the relationship between world regions' personality structure and the structure of those face impressions in those regions. For instance, if a perceiver observes that aggressive people tend be intelligent, they will conceptually associate those traits as co-occurring; in turn, that perceiver may use similar facial appearance to judge whether targets are aggressive or intelligent. In Study 2, we tested the possibility that regional variability in personality structure is reflected in the structure of face impressions, which may be partly explained by regional perceivers' conceptual trait structure.

Method

Previous research with U.S. samples has shown that U.S. conceptual trait associations predict the structure of U.S. face-based trait judgments (Stolier et al., 2018, 2020). We used these previous U.S. data on conceptual trait associations and face-based trait judgments in tandem with the publicly available data set of NEO-PI personality across world regions used in Study 1 (which includes the United States). Using RSA, we were then able to test whether the similarity in a given world region's personality structure to that of the United States can predict how similarly that region's personality structure also resembles the structure of U.S. conceptual beliefs and U.S. face-based judgments. For instance, if Australia's personality structure is more similar to U.S. personality structure than is Syria's, we would expect Australia's personality structure to also more closely resemble the structure of U.S. conceptual trait beliefs and U.S. face-based judgments than Syria's personality structure. Thus, although widespread cross-regional data such as NEO-PI data and face-judgment data (used in Study 1) are not available for conceptual trait associations, using cross-regional NEO-PI and face-judgment data in tandem with full data from the United States (NEO-PI, face-judgment, and conceptual trait data) provided a valuable opportunity to test our hypothesis regarding the intermediary role of conceptual trait associations.

Participants. For the conceptual trait associations, we used published data from 115 U.S. participants (age: M = 35.38 years, SD = 10.47; 47.83% male, 50.43% female, 1.74% declined to report gender or reported another gender; Stolier et al., 2020, Study 1). For the face-based trait judgments, we used data from 482 U.S. participants (age: M = 35.51 years, SD = 12.30; 41.29% male, 58.30% female, 0.42% declined to report gender or reported another gender). To conduct an additional replication, we used an additional sample of face-based trait judgments from 496 participants (age: M = 30.31 years, SD = 6.74; 47.78% male, 51.81% female, 0.40% declined to report gender or reported another gender; Stolier et al., 2020, Studies 1 and 2). All data were taken from the study by Stolier et al. (2020). The research protocols used for data collection were approved by the University Committee on Activities Involving Human Subjects at New York University.

For the NEO-PI personality data, we used the same personality data as in Study 1 of participants from different world regions (Johnson, 2014). Because here we were not constrained by the subset of regions also available in the face-judgment data set used in Study 1, for the present study what remained was a sample of 307,136 personality respondents across 232 regions, including the United States (age: M = 25.19 years, SD = 10.00; 39.74% male, 60.26% female). See Figure S1b in the Supplemental Material for the complete list and geographic locations of the regions. All participant samples were convenience samples. We used all available participants' data, and we did not predetermine the sample size.

NEO-PI personality data. For the stimuli and procedure used for the personality data collection, see Study 1.

Conceptual-trait-association data. For the trait-association rating task, participants were asked to provide conceptual similarity ratings for all pairwise combinations of 15 personality traits: adventurous, angry, anxious, assertive, cautious, cheerful, cooperative, depressed, dutiful, emotional, friendly, intellectual, self-disciplined, sympathetic, and trustworthy. The 15 traits represented 15 NEO-PI facets (three facets representing each of the five NEO-PI factors). These 15 representative facets of the total 30 were found to be able to explain various domains of social perception, including representations of social groups and face impressions of strangers (Stolier et al., 2020). For example, for the pair of "adventurous" and "assertive," participants were asked, "How likely is an [adventurous] person to be [assertive]?" Responses were made on a 7-point scale (1 = *not at all likely*, 7 = *very likely*). Participants evaluated faces on the same 15 personality traits. Each participant rated the degree of association across all 105 unique trait pairs, and all traits were presented twice to capture the association bidirectionally (e.g., how likely an adventurous person is to be assertive, and how likely an assertive person is to be adventurous).

Details of the personality-trait-association rating procedure can be found in Stolier et al. (2020).

Face-impressions data. For face-based personality trait judgments, participants judged 90 target faces on the same 15 personality traits (15 NEO-PI facets) used in the conceptual-trait-association task. All images were of an identical race and gender (White male) and taken from the Chicago Face Database (Ma et al., 2015). Independent groups of participants (n = 25–30 for each group) were assigned to each of the 15 traits, and thus participants rated faces only on a single trait. Participants provided 7-point ratings (e.g., 1 = *not at all [adventurous]*, 7 = *very [adventurous]*). Details of the face-judgment procedure can be found in the study by Stolier et al. (2020).

Language covariates. As in Study 1, we repeated all analyses while including covariates related to language use. Here, all language-distance metrics captured the distance between language use of any given region and the United States. The official language of each region was considered its primary language; if English was one of a region's multiple official languages, English was considered the primary language. The ASJP language distance was used to index language dissimilarity between a region's primary language and English, and the difference between a region's TOEFL score and the United States' TOEFL score was used to index the difference in a region's facility with English relative to the United States' facility with English. An EPI difference score was not included in the models because EPI is not measured in regions in which English is widely spoken as a first language, including the United States. Among the 232 world regions considered, all 232 regions had their primary languages (69 languages in total) in the ASJP database available (100%), and 172 regions had regional TOEFL scores available (74.46%).

Ethnic-diversity covariate. To consider regional differences in ethnic diversity, as in Study 1, we included an ethnic-fractionalization HHI difference score (between a region's HHI and the United States' HHI) as a covariate. The HHI was available for 166 regions (71.55% of all regions).

Analytic approach. Because participants in the conceptual-trait-association and face-judgment tasks evaluated the identical 15 personality traits as the 15 NEO-PI facets, data could be linked across NEO-PI personality data, conceptual-trait data, and face-judgment data at the level of the same 15 personality traits. Unlike Study 1's trait-level RSA (but similar to Study 1's region-level RSA), analyses in Study 2 did not require a trait-level correspondence between the different data sources. Thus, personality–face mappings were not necessary for Study 2. For each of the 231 non-U.S. regions, we calculated three measures, each a correlation between the non-U.S. region's personality structure and the U.S.'s personality structure (U.S.-to-region personality correlation), the U.S.'s conceptual-trait structure (U.S.-to-region conceptual-trait correlation), and the U.S.'s face-judgment structure (U.S.-to-region face-judgment correlation). For each of the three measures, there were 231 final values corresponding to the 231 world regions.

The first measure, the U.S.-to-region personality correlation, was calculated as the Pearson correlation between the 300-item NEO-PI trait scores of the United States and the 300-item NEO-PI trait scores of the non-U.S. region. This measure thus represents how similar the personality structure is between the United States and any given world region.

The second measure, the U.S.-to-region conceptual-trait correlation, was computed using RSA. We first created two 15 × 15 dissimilarity matrices, one for the United States' conceptual-trait data and one for the non-U.S. region's NEO-PI data, with cells reflecting the dissimilarity (euclidean distance) between all pairwise combinations of the 15 personality traits. Cells of the U.S. conceptual-trait 15 × 15 dissimilarity matrix reflected U.S. participants' conceptual beliefs that any given pair of traits tends to co-occur in other people; cells of the non-U.S. region's NEO-PI 15 × 15 dissimilarity matrix reflected the extent to which that same pair of traits tends to actually co-occur in other people's personalities in the region. We vectorized the 105 unique values in the two dissimilarity matrices and assessed their Pearson correlation. This U.S.-to-region conceptual-trait correlation thus represents the correspondence between a given non-U.S. region's NEO-PI personality structure and the U.S.'s conceptual-trait structure.

The third measure, the U.S.-to-region face-judgment correlation, was also computed as RSA. Two 15 × 15 dissimilarity matrices were created, one for the United States' face-judgment data and one for the non-U.S. region's NEO-PI data, with cells reflecting the dissimilarity (euclidean distance) between all pairwise combinations of the 15 personality traits. Cells of the U.S. face-judgment 15 × 15 dissimilarity matrix reflected U.S. participants' tendencies to judge two personality traits similarly in response to the same faces; cells of the non-U.S. region's NEO-PI 15 × 15 dissimilarity matrix reflected the extent to which that same pair of traits tends to actually co-occur in other people's personalities in the region. We vectorized the 105 unique values in the two dissimilarity matrices and assessed their Pearson correlation. This U.S.-to-region face-judgment correlation thereby represents the correspondence between a given non-U.S. region's NEO-PI personality structure and the United States' face-impressions structure.

Mediation analyses were used to test the intermediary role (i.e., indirect effect) of conceptual trait associations. As in Study 1, we also reran all analyses after including covariates related to language use and ethnic diversity. We also conducted an additional corroborating analysis; the larger number of regions analyzed in Study 2 allowed us to conduct multilevel GEE regression analyses that clustered by primary language. If the relationships of interest persisted even within clusters of regions with the same primary language (e.g., within the 75 English-speaking regions, within the 26 French-speaking regions, within the 20 Arabic-speaking regions, and within the 19 Spanish-speaking regions), this would help to cement the evidence that the effects of NEO-PI structure via conceptual-trait structure are not confounded by language.

Results

Using RSA, previous reports of the U.S. data have shown that the structure of U.S. NEO-PI personality predicts the structure of U.S. conceptual-trait associations (Spearman's $\rho = .77$, 95% CI = [.68, .84], $p < .001$; Study 7, Stolier et al., 2020), which in turn predicts the structure of U.S. face impressions (Spearman's $\rho = .80$, 95% CI = [.71, .86], $p < .001$; Study 1, Stolier et al., 2020). This previous result suggests that actual personality in the U.S. environment may shape the structure of U.S. perceivers' conceptual understanding of personality, which in turn sets the stage for their judgments of others' faces. Study 1 provided evidence that regional variability in personality structure relates to regional variability in the structure of face impressions. Our analyses here focused on helping to explain this regional association described in Study 1 by way of conceptual trait associations, using international data across 232 world regions (including the United States) in terms of their NEO-PI personality structure (as in Study 1) together with United States–only data on face impressions and conceptual-trait structures. To the extent that any given world region is more similar to the United States in terms of NEO-PI personality structure, that region's NEO-PI structure should be able to more strongly predict the United States' conceptual trait structure, and, in turn, face-judgment structure (relative to other regions less similar to the United States in terms of personality structure).

The three variables of interest were (a) U.S.-to-region personality correlation (correlation between regional NEO-PI structure and U.S. NEO-PI structure); (b) U.S.-to-region conceptual-trait correlation (correlation between regional NEO-PI structure and U.S. conceptual-trait structure); and (c) U.S.-to-region face-judgment correlation (correlation between regional NEO-PI structure and U.S. face-judgment structure). The three variables (r coefficients) for the 231 world regions were assessed with Spearman correlation analyses, which revealed that they were all positively correlated (Spearman's ρs = .32–.71, ps < .001). Thus, if a given region (e.g., Australia) was more similar in personality structure to the United States, then that region's personality structure was better able to predict the United States' conceptual-trait structure and face-impressions structure.

To test the possibility that U.S.-to-region conceptual-trait correlations (mediator) may partly explain the relationship between U.S.-to-region personality correlations (independent variable) and U.S.-to-region face-judgment correlations (dependent variable), we conducted a mediation analysis. As expected, given the correlational analyses above, the independent variable was strongly related to both the dependent variable, $b = 0.39$, $SE = 0.06$, 95% CI = [0.28, 0.51], $t(229) = 6.49$, $p < .001$, and the mediator, $b = 0.32$, $SE = 0.06$, 95% CI = [0.20, 0.44], $t(229) = 5.15$, $p < .001$. Further, the relationship between the mediator and the dependent variable remained significant even after we statistically controlled for the independent variable, $b = 0.65$, $SE = 0.05$, 95% CI = [0.55, 0.74], $t(228) = 13.56$, $p < .001$. Most importantly, bootstrapping analyses demonstrated a significant indirect effect, by which U.S.-to-region conceptual-trait correlations (mediator) partly explained the relationship between U.S.-to-region personality correlations (independent variable) and U.S.-to-region face-judgment correlations (dependent variable), $b = 0.21$, 95% CI = [0.13, 0.29], $p < .001$ (Fig. 3b; see also Table S6a in the Supplemental Material).

To demonstrate robustness and generalizability of the effects, we reran analyses used a complementary dependent variable. Rather than being asked to judge faces directly on the basis of 15 trait adjectives (e.g., "How likely is this person to be [adventurous]?"), a separate group of U.S. participants was asked to judge faces using phrase descriptions as stand-ins for the 15 traits (e.g., "How likely is this person to [enjoy visiting new places]?"). A full list of phrase descriptions associated with traits is available in Table S4 in the Supplemental Material; data were taken from the study by Stolier et al. (2020). We again observed strong positive relationships between the independent and the dependent variable, $b = 0.30$, $SE = 0.06$, 95% CI = [0.18, 0.43], $t(229) = 4.79$, $p < .001$, the independent variable and the mediator, $b = 0.32$, $SE = 0.06$, 95% CI = [0.20, 0.44], $t(229) = 5.15$, $p < .001$, and the mediator and the dependent variable, $b = 0.75$, $SE = 0.04$, 95% CI = [0.66, 0.84], $t(228) = 16.86$, $p < .001$, as well as a significant indirect effect, $b = 0.24$, 95% CI = [0.14, 0.34], $p < .001$ (see Fig. S3 and Table S6b in the Supplemental Material). Note that the relationship between the independent variable and the mediator does

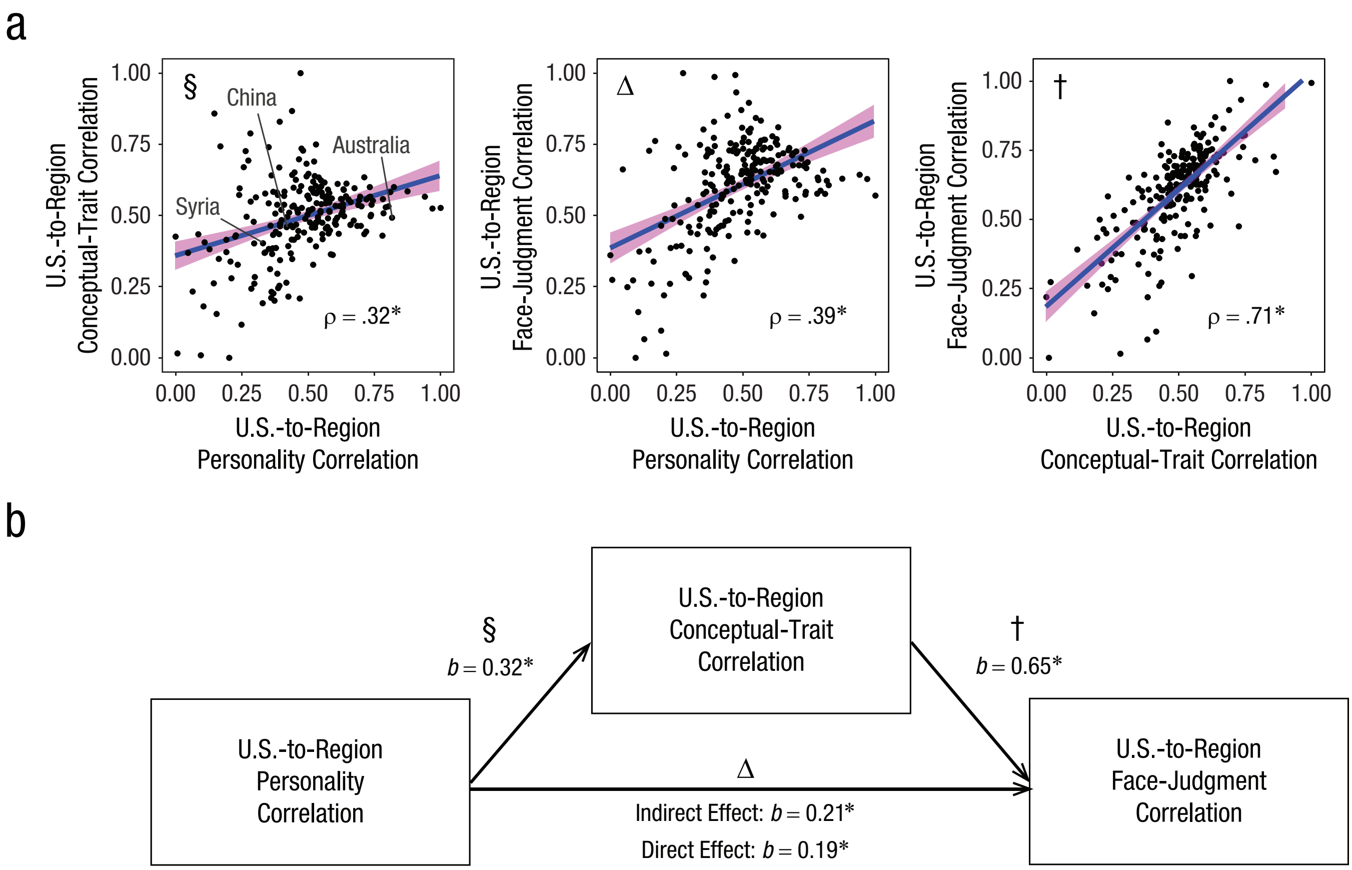

Fig. 3. Results of Study 2. The scatterplots (a) show the relationships between three measures, each a correlation between the non-U.S. region's personality structure and the U.S.'s personality structure (U.S.-to-region personality correlation), the U.S.'s conceptual-trait structure (U.S.-to-region conceptual-trait correlation), and the U.S.'s face-judgment structure (U.S.-to-region face-judgment correlation). The blue line in each plot represents the best-fitting linear regression, and the shaded area represents the 95% confidence interval. Dots indicate each of the 231 non-U.S. regions. For illustrative purposes, both x and y coordinates are displayed using unranked Pearson r coefficients; actual analyses were run using rank-ordered r coefficients (i.e., by running three Spearman-rank correlations across three measures, which were in the form of rank-ordered values of Pearson rs). The model (b) shows the influence of U.S.-to-region personality correlation on U.S.-to-region face-judgment correlation, as mediated by U.S.-to-region conceptual-trait correlation. Asterisks in (a) indicate significant correlations and in (b) indicate significant paths ($p < .001$).

not involve the dependent variable; thus, this result was identical regardless of whether trait words or phrases were used for the dependent variable.

The results were also robust to the inclusion of the language-use and ethnic-diversity covariates. We again observed positive relationships between the independent variable and the dependent variable, $b = 0.32$, $SE = 0.09$, 95% CI = [0.14, 0.51], $t(134) = 3.42$, $p = .001$, the independent variable and the mediator, $b = 0.32$, $SE = 0.09$, 95% CI = [0.14, 0.51], $t(134) = 3.47$, $p = .001$, and, critically, the mediator and the dependent variable when we controlled for the independent variable, $b = 0.70$, $SE = 0.06$, 95% CI = [0.58, 0.82], $t(133) = 11.06$, $p < .001$, as well as a significant indirect effect, $b = 0.22$, 95% CI = [0.10, 0.35], $p < .001$. We reran the same analysis using the complementary dependent variable (i.e., face-trait ratings derived from trait phrases), and the results were unchanged: We again observed a significant indirect effect, $b = 0.24$, 95% CI = [0.11, 0.39], $p < .001$. See Tables S5 and S6 in the Supplemental Material for full statistics.

The larger number of regions available in Study 2 afforded an additional corroborating analysis that clustered regions by primary language using multilevel GEE regression. We considered only sets of same-language regions with sufficient size (≥ 10 regions per language). This resulted in clusters of 75 English-, 26 French-, 20 Arabic-, and 19 Spanish-speaking regions (140 regions in total, 61% of all 231 non-U.S. regions). If the relationships of interest persisted even within clusters of regions with the same primary language, that would represent strong evidence that the effects of NEO-PI structure via conceptual-trait structure are not confounded by language. To further control for potential confounding effects of language use and ethnic diversity within the same-language groups, we included the same language-use and ethnic-diversity covariates.

Clustering by language across the 140 regions revealed virtually identical results. Whether we used face-trait ratings derived from trait words or trait phrases, U.S.-to-region personality correlations predicted U.S.-to-region face-judgment correlations (trait words: b = 0.28, SE = 0.05, 95% CI = [0.18, 0.38], Z = 5.40, $p < .001$; trait phrases: b = 0.39, SE = 0.09, 95% CI = [0.21, 0.58], Z = 4.25, $p < .001$) and predicted U.S.-to-region conceptual trait correlations (b = 0.34, SE = 0.13, 95% CI = [0.01, 0.59], Z = 2.72, p = .007). U.S.-to-region conceptual trait correlations also predicted U.S.-to-region face-judgment correlations (trait words: b = 0.69, SE = 0.03, 95% CI = [0.62, 0.75], Z = 20.66, $p < .001$; trait phrases: b = 0.71, SE = 0.11, 95% CI = [0.49, 0.91], Z = 6.50, $p < .001$). See Table S7 in the Supplemental Material for full statistics. Mediation analysis and estimates of the indirect effect are not possible with multilevel GEE regression.

In sum, these results suggest that when a region's personality structure was similar to that of the United States (e.g., Australia), that region's personality structure could more strongly predict the structure of face-based trait judgments in the United States (better than could regions whose personality structure was dissimilar to that of the United States, e.g., Syria). Importantly, this relationship was partly explained by how well that region's personality structure could predict conceptual trait associations in the United States, even when analyses controlled for language use and ethnic diversity. The results held regardless of whether participants were asked to judge trait adjectives (e.g., "adventurous") or to judge phrase descriptions (e.g., "enjoy visiting new places"), which alleviates concerns that correspondence between face impressions, trait concepts, and actual personality structure may be solely due to semantic confounds (i.e., using the same adjectives in all tasks).

Discussion

The present results replicate those of Study 1, showing that a region's face impressions reflect its personality structure. Furthermore, the findings implicate conceptual trait associations as playing an intermediary role in the relationship between regional variability in personality structure and regional variability in face-impressions structure.

General Discussion

In two studies, we found that the actual personalities of people in a region are related to how individuals in that region judge others' traits from faces. For example, in a region where people were more likely to be simultaneously aggressive and intelligent, people were more likely to judge a person with a face appearing more aggressive as more intelligent (Study 1). Moreover, the personality structure of regions that were more similar to the United States in personality better predicted U.S. conceptual-trait structure and U.S. face-impressions structure (Study 2). These effects generalized across different ways of assessing face impressions (adjectives or phrases), alleviating the concern of semantic confounds. Together, the findings suggest that people form face-based inferences of others' personalities on the basis of a conceptual understanding of personality that they learn from their regional environment.

The fact that people's face impressions vary depending on the social environment is consistent with evidence for the role of learning in face impressions (Dotsch et al., 2016; Stolier et al., 2020; Sutherland et al., 2020). The role of conceptual associations in guiding face impressions extends previous studies (Stolier et al., 2018, 2020) by implicating these associations as a mechanism by which region-specific social experience can affect face impressions. Because personality structure (McCrae & Costa, 1997) and face-impressions structure show a general consistency across cultures (Todorov & Oh, 2021), variations in these structures have often been overlooked; departures from a universal dimensional structure have been described as statistical noise. Our findings bridge variability in personality and variability in face impressions, demonstrating that this noise may contain information about person perception. Regional differences in personality have been suggested to result from various regional factors, such as culture (McCrae & Terracciano, 2005) and socioecological complexity (Lukaszewski et al., 2017). Future research could examine how these factors affect not only the actual personalities of local residents but also how those residents think about personality and judge personality in others.

Our approach was correlational, which afforded a comprehensive assessment across a large number of regions but which limited the ability to make causal claims. Although we propose that conceptual associations in the form of lay theories of personality serve as a causal mechanism linking personality in the environment to face impressions, this possibility was not directly tested. The potential roles of other intermediary factors, such as cultural differences in basic face processing (e.g., Caldara, 2017; Hills & Pake, 2013) could be examined in future research. Another limitation is that English-only questionnaires were used to obtain the personality data. Prior work has shown that personality structure is highly similar when the NEO-PI questionnaire is administered in English as opposed to a respondent's native language (McCrae, 2001), and any respondent in our data sets whose data suggested poor English comprehension or confusion was excluded (Johnson, 2005, 2014). Nevertheless, we comprehensively controlled for the potential confounding role of language. Using multiple measures of regions' English proficiency and the linguistic similarity between any given region's primary languages, as well as corroborating analyses that tested our effects within regions of the same language, we found no evidence that language confounded the results. However, future research could collect personality data in participants' native languages to further investigate the potential interplay of regional language and personality in shaping face-impressions structure.

It is important to recognize that the present results cannot directly speak to questions on the accuracy of face impressions. Our findings can speak only to how traits are judged from faces, not how they may manifest or be expressed on people's faces. Even if people learned an accurate trait structure, it could help them accurately infer a person's personality traits only when they already possessed accurate information about another trait (which covaries with the trait in question). Thus, accurately learning the structure of personality traits in the social environment need not imply that perceivers can accurately intuit specific traits in others. Future research could explore these questions directly.

In sum, the current results suggest that perceivers use the actual personality structure learned from their social environment to form lay theories about personality, and these beliefs in turn support the structure of perceivers' face impressions. The findings call for a greater focus on the regional and cultural specificity of face impressions and the role of social experience in how we infer personality from facial appearance.

Transparency

Action Editor: Paul Jose
Editor: Patricia J. Bauer
Author Contributions
All the authors developed the study concept and contributed to the study design. D. Oh and J. D. Martin analyzed and interpreted the data under the supervision of J. B. Freeman. D. Oh drafted the manuscript, and all the authors provided critical revisions. All the authors approved the final version of the manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported in part by Research Grant No. BCS-1654731 to J. B. Freeman.
Open Practices
All data and analysis scripts have been made publicly available via OSF and can be accessed at https://osf.io/yazn6/. The design and analysis plans for the studies were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

DongWon Oh https://orcid.org/0000-0002-2105-3756
Jonathan B. Freeman https://orcid.org/0000-0002-2061-8460

Acknowledgments

We thank Rick Dale, Karina Tachihara, and Seongyong Lee for their insight regarding linguistic measures.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211072814

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, *8*(2), 155–194.

Allik, J., Church, A. T., Ortiz, F. A., Rossier, J., Hřebíčková, M., de Fruyt, F., Realo, A., & McCrae, R. R. (2017). Mean profiles of the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, *48*(3), 402–420. https://doi.org/10.1177/0022022117692100

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Languages evolve in punctuational bursts. *Science*, *319*(5863), Article 588. https://doi.org/10.1126/science.1149683

Bentz, C., Dediu, D., Verkerk, A., & Jager, G. (2018). The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour*, *2*(11), 816–821. https://doi.org/10.1038/s41562-018-0457-6

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences*, *69*, 56–61. https://doi.org/10.1016/j.paid.2014.05.012

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*(3), 249–255.

Church, A. T., & Katigbak, M. S. (2002). The five-factor model in the Philippines. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 129–154). Springer.

Collova, J. R., Sutherland, C. A., & Rhodes, G. (2019). 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*, *117*(5), 900–924.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO Personality Inventory-Revised (NEO-PI-R)*. Psychological Assessment Resources.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, *1*(1), Article 0001. https://doi.org/10.1038/s41562-016-0001

Educational Testing Service. (2021). *TOEFL iBT: Test and score data summary 2020*.

Education First. (2020). *EF English Proficiency Index: Global ranking of countries and regions*. https://www.ef.com/wwen/epi

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion*, *2*(1), 75–84.

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

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

Gülgöz, S. (2002). Five-factor model and NEO-PI-R in Turkey. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 175–196). Springer.

Han, C., Wang, H., Hahn, A. C., Fisher, C. I., Kandrik, M., Fasolt, V., Morrison, D. K., Lee, A. J., Holzleitner, I. J., DeBruine, L. M., & Jones, B. C. (2018). Cultural differences in preferences for facial coloration. *Evolution and Human Behavior*, *39*(2), 154–159. https://doi.org/10.1016/j.evolhumbehav.2017.11.005

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*(3), 586–597. https://doi.org/10.1016/j.cognition.2013.08.012

Hirschman, A. O. (1945). *National power and the structure of foreign trade*. University of California Press.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008a, June). *Advances in automated language classification* [Paper presentation]. 3rd Workshop on Quantitative Investigations in Theoretical Linguistics (QITL3), Helsinki, Finland.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008b). Explorations in automated language classification. *Folia Linguistica*, *42*(2), 331–354.

Holzleitner, I. J., & Perrett, D. I. (2017). Women's preferences for men's facial masculinity: Trade-off accounts revisited. *Adaptive Human Behavior and Physiology*, *274*(3), 1–17.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 199–209. https://doi.org/10.1037/0096-1523.32.2.199

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, USA*, *109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Johnson, J. A. (2014). 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.

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., Mburu, G., . . . Coles, N. A. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*, *5*, 159–169. https://doi.org/10.1038/s41562-020-01007-2

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

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

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterizing trait attributions from faces. *Nature Communications*, *12*, Article 5168. https://doi.org/10.1038/s41467-021-25500-y

Lukaszewski, A. W., Gurven, M., von Rueden, C. R., & Schmitt, D. P. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science*, *8*(8), 943–952.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Martinez, J. E., Funk, F., & Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to stimulus evaluations. *Behavior Research Methods*, *52*, 1428–1444.

McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons. *Journal of Personality*, *69*(6), 819–846.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 105–125). Springer.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509–516.

McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*(3), 407–425.

McCrae, R. R., Yik, M. S., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. *Journal of Personality and Social Psychology*, *74*(4), 1041–1055.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*(2), 323–342. https://doi.org/10.1037/xge0000638

Oh, D., Grant-Villegas, N., & Todorov, A. (2020). The eye wants what the heart wants: Females' preference in male faces are related to partner personality preference. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(11), 1328–1343. https://doi.org/10.1037/xhp0000858

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

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720. https://doi.org/10.1038/nature06176

Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T., Jr. (2002). The applicability of the five-factor model in a sub-Saharan culture. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 155–173). Springer.

Piedmont, R. L., & Chae, J.-H. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO PI-R for Koreans. *Journal of Cross-Cultural Psychology*, *28*(2), 131–155.

Pompei, S., Loreto, V., & Tria, F. (2011). On the accuracy of language trees. *PLOS ONE*, *6*(6), Article e20109. https://doi.org/10.1371/journal.pone.0020109

Rolland, J.-P. (2002). The cross-cultural generalizability of the five-factor model of personality. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 7–28). Springer.

Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kamble, S., Muluk, H., Masuda, T., & Niedenthal, P. M. (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences, USA*, *112*(19), E2429–E2436. https://doi.org/10.1073/pnas.1413661112

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, *79*(5), 294–309.

Simakhodskaya, Z. (2000, August). *Russian revised NEO-PI-R: Concordant validity and relationship to acculturation* [Paper presentation]. The 108th Annual Convention of the American Psychological Association, Washington, D.C.

Søren, W., Holman, E. W., & Brown, C. H. (Eds.) (2020). *The ASJP database* (Version 19.1). https://asjp.clld.org

Soto, J. A., & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage. *Emotion*, *9*(6), 874–884.

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

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences, USA*, *115*(37), 9210–9215. https://doi.org/10.1073/pnas.1807222115

Sutherland, C. A. M., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences, USA*, *117*(19), 10218–10224. https://doi.org/10.1073/pnas.1920131117

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186–208.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 63, pp. 189–245). Academic Press.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.

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

Xie, S. Y., Flake, J. K., & Hehman, E. (2018). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. https://doi.org/10.1037/pspi0000160

Zhang, L., Holzleitner, I. J., Lee, A. J., Wang, H., Han, C., Fasolt, V., DeBruine, L. M., & Jones, B. C. (2019). A data-driven test for cross-cultural differences in face preferences. *Perception*, *48*(6), 487–499. https://doi.org/10.1177/0301006619849382