Running Head: ICC

THIS PAPER WAS ACCEPTED FOR PUBLICATION IN THE JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY ON APRIL 10 2017. THIS IS THE NON COPY EDITED, IN PRESS VERSION

The Unique Contributions of Perceiver and Target Characteristics in Person Perception

Eric Hehman[1], Clare A. M. Sutherland[2], Jessica K. Flake[3], Michael L. Slepian[4]

[1]Department of Psychology, Ryerson University, Toronto, ON, M5B 2K3
[2]ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, Crawley 6009, WA, Australia
[3]Department of Psychology, York University, Toronto, ON, M3J 1P3
[4]Department of Management, Columbia University, New York, NY, 10027

Corresponding Author:
Eric Hehman, Department of Psychology, Ryerson University, E-mail: eric.hehman@psych.ryerson.ca
Author Contributions:
All authors designed the studies. E.H. aggregated the data. E.H. analyzed the data. All authors wrote the manuscript.
Research Support:
This research was partially supported by a SSHRC Institutional Grant and SSHRC Insight Development Grant (430-2016-00094) to EH and postdoctoral research support from the Australian Research Council Centre of Excellence in Cognition and its Disorders, University of Western Australia (CE110001021) and an Australian Research Council Discovery Project Grant (DP170104602) to CS.
Acknowledgements:
We thank Vito Adamo for his help in compiling the data.

Word count: 12,438

Abstract

Models of person perception have long asserted that our impressions of others are guided by characteristics of both the target and perceiver. However, research has not yet quantified to what extent perceivers and targets contribute to different impressions. This quantification is theoretically critical, as it addresses how much an impression arises from “our minds” vs. “others’ faces”. Here, we apply cross-classified random effects models to address this fundamental question in social cognition, using approximately 700,000 ratings of faces. With this approach, we demonstrate that 1) different trait impressions have unique causal processes; meaning that some impressions are largely informed by perceiver-level characteristics whereas others are driven more by physical target-level characteristics, 2) modeling of perceiver- and target-variance in impressions informs fundamental models of social perception, 3) perceiver × target interactions explain a substantial portion of variance in impressions, 4) greater emotional intensity in stimuli decreases the influence of the perceiver, and 5) more variable, naturalistic stimuli increases variation across perceivers. Important overarching patterns emerged. Broadly, traits and dimensions representing inferences of character (e.g., dominance) are driven more by perceiver characteristics than those representing appearance-based appraisals (e.g., youthful-attractiveness). Moreover, inferences made of more ambiguous traits (e.g., creative) or displays (e.g., faces with less extreme emotions, less-controlled stimuli) are similarly driven more by perceiver than target characteristics. Together, results highlight the large role that perceiver and target variability play in trait impressions, and develop a new topography of trait impressions that considers the source of the impression.

Keywords: impression formation, person perception, face perception, multilevel modeling

The Unique Contributions of Perceiver and Target Characteristics in Person Perception

To what extent are our perceptions subjective? This fundamental question, considered by philosophers for centuries, has over time transformed into an idea at the very core of modern social cognition. To what extent do our impressions of others arise from two distinct sources: the target and the perceiver? Many models of person perception have been constructed to explain how the physical characteristics of the individuals being observed lead to impressions. The perceiver, however, is no blank canvas onto which the targets project these impressions. Rather, perceivers interpret what they observe, and final impressions are additionally influenced by a host of perceiver-level factors.

Beginning with the cognitive revolution, multiple social and cognitive models have described these two sources of information as influencing impressions of people (Bruce & Young, 1986; Brunswik, 1952; Correll, Hudson, Guillermo, & Earls, 2016; Haxby, Hoffman, & Gobbini, 2000; Kenny & Albright, 1987; Kunda et al., 1996; Neuberg & Fiske, 1987; West & Kenny, 2011). Informed by recent insights into brain function and cognitive processes, recent models have grown in complexity (e.g., detailing dynamic temporal processes; Freeman & Ambady, 2011). Surprisingly, these social-cognitive models have yet to specify the *extent* to which perceiver- and target-level characteristics influence impressions, and how these inputs may vary across different trait impressions and contexts. Understanding the relative contribution of these two sources of inputs to impression formation is paramount to understanding the very nature of how perceivers form first impressions. To provide an analogy, just as one cannot fully understand the etiology of a disease without understanding the relative contributions of genetics and experience (i.e., nature and nurture), one cannot fully understand a formed impression without understanding the extent to which it is driven by perceiver- and target-level characteristics, and their interaction. Without detailing *the extent* to which perceivers contribute to impressions of others the process of impression formation itself remains obscure. In short, to understand the extent to which perceiver- and target-level factors respectively influence our impressions of others is to better understand the processes by which perceivers evaluate others.

To this end, the goal of the current work was to address this question at the very core of social cognition: “To what extent do first impressions arise from the perceiver versus the target?” We do so by applying recently developed statistical methods to ~700,000 trait ratings from faces, from ~7,000 participants rating ~3,000 stimuli. We focus on impressions of faces, as they are critical for human social perception (Webster & Macleod, 2011), attended to early in development (Sugden, Mohamed-Ali, & Moulson, 2014), provide a wealth of cues to first impressions (Bruce & Young, 2011), and are reasonably well theoretically understood (Rhodes, 2006; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2014; Zebrowitz & Montepare, 2005). Therefore, we quantify the relative unique contributions of the perceiver and target for a wide variety of important impressions from faces. We note, however, that the principles outlined here apply to any facet of social perception.

With cross-classified multilevel models (described further below), we demonstrate that 1) different trait impressions have unique causal processes; meaning that some impressions are largely informed by perceiver-level characteristics whereas others are driven more by physical target-level characteristics, 2) modeling of perceiver- and target-variance in impressions informs fundamental models of social perception, 3) the unique interplay between characteristics of perceivers and targets explains a substantial portion of variance in impressions, 4) increasing emotional intensity in the target stimuli decreases the influence of perceiver-level characteristics, and 5) more variable, naturalistic stimuli also increases variation across perceivers.

Quantifying perceiver and target contributions develops a theoretically richer and nuanced understanding of an impression than when perceiver and target contributions are conflated. In addition to addressing substantive questions within the domain of person perception, we also aim to illustrate the utility of cross-classified multilevel models by providing researchers with the tools to use these models in their own research (see the supplementary materials for annotated R code). As such, this paper meets two ends: providing a better theoretical understanding of the contributions of the person and the target in impression formation, as well as a demonstration of how this underutilized statistical approach can be implemented to inform theoretical models in general.

Target Contributions to Impressions

It is intuitive that a target's facial features influence perceiver impressions of that target, and research over the past several decades has contributed to an increased understanding of which cues are involved (for review, see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2014). In the initial moments after encountering someone, features of the face are used to help identity to which social categories an individual might belong (Freeman & Ambady, 2011; Freeman, Pauker, Apfelbaum, & Ambady, 2010; Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014; Kubota & Ito, 2007), or what emotion they may be experiencing (Adams, Nelson, Soto, Hess, & Kleck, 2012; Bruce & Young, 1986; Darwin, 1872; Ekman & Friesen, 1971). Moreover, slight resemblances to emotional expressions, either through natural variations in facial structure or temporary muscle contractions, are overgeneralized to corresponding trait inferences (Adams, Garrido, Albohn, Hess, & Kleck, 2016; Oosterhof & Todorov, 2009; Said, Sebe, & Todorov, 2009; Secord & Bevan, 1956; Zebrowitz, Kikuchi, & Fellous, 2007, 2010). For instance, a person with naturally down-turned brows can be evaluated as less friendly due to similarities with angry emotional expressions, and individuals with rounder faces and larger eyes are evaluated as more innocent and warm due to shared structural similarity with babies' faces (Zebrowitz & Montepare, 1992).

Because emotional resemblance is largely based on a face's underlying musculature, these emotional expressions are fluid and dynamic (Hehman, Flake, & Freeman, 2015; Sutherland, Young, & Rhodes, 2016; Todorov & Porter, 2014), but relatively static morphological features of the face can additionally influence perceptions. For instance, the width of a face relative to its height has been linked to perceptions of physical aggression and strength (Carré, McCormick, & Mondloch, 2009; Carré, Morrissey, Mondloch, & McCormick, 2010; Hehman, Leitner, Deegan, & Gaertner, 2015; Hehman, Leitner, & Gaertner, 2013). The symmetry (Rhodes, 2006; Rhodes et al., 2001) and skin coloration of a face (Re, Whitehead, Xiao, & Perrett, 2011; Stephen, Law Smith, Stirrat, & Perrett, 2009) are linked to attractiveness, and facial height has been associated with perceptions of leadership ability (Re, Hunter, et al., 2013; Re, DeBruine, Jones, & Perrett, 2013).

While perceivers are apparently inaccurate in forming some impressions from appearance, such as perceptions of trustworthiness (Olivola & Todorov, 2010; Rule, Krendl, Ivcevic, & Ambady, 2013; but see Slepian & Ames, 2016), there may be a kernel of truth to other perceptions, such as extraversion, prejudice, or sexual unfaithfulness (Ambady & Rosenthal, 1992; Carney, Colvin, & Hall, 2007; Funder, 2012; Hehman, Leitner, Deegan, & Gaertner, 2013; Rhodes, Morley, & Simmons, 2013). A higher degree of accuracy may indicate that there is a greater "signal" in faces for some traits than others, and thus that target-level factors are contributing to the final rating to a greater extent. Regardless of accuracy, it is clear that humans are very sensitive to diverse yet often subtle facial variation, from which robust inferences of target characteristics are inferred. Without an understanding of how perceiver- and target-level factors influence these impressions, we cannot begin to build a broad model of impression formation, nor identify how different trait ratings compare and contrast from each other.

Perceiver Contributions to Impressions

While there may exist signals in the face to inform accurate judgments, the perceiver is not passive in the process of forming impressions. A host of individual differences might influence impressions. For example, temporary cognitive states can alter perceptions. When perceivers are feeling threatened they consistently evaluate targets as larger and more dangerous (Fessler & Holbrook, 2013a, 2013b). In the present work "perceiver characteristics" captures any of the ways in which perceiver factors might exert a consistent influence on impressions.

In addition, characteristics of the perceiver may uniquely interact with characteristics of the target in determining particular impressions. For instance, racial prejudice facilitates interpreting facial features as hostile on other-race, but not own-race, faces (Hugenberg & Bodenhausen, 2003). Idiosyncratic experiences, such as how much a target resembles people who are familiar to the perceiver, also influence perceptions of the target (DeBruine, 2002; DeBruine, Jones, Little, & Perrett, 2008; Verosky & Todorov, 2013). Different experiences across one's lifetime such as quantity of contact with members of different social groups (Freeman, Pauker, & Sanchez, 2016) or majority/minority status (Hehman et al., 2012; Verkuyten, 2005) influence how individuals of different groups are perceived. We refer to impressions that are jointly determined by perceiver and target characteristics as perceiver × target interactions.

Linking Theory with Statistical Models

Crucially, the relative contributions of perceiver- and target-level characteristics for different trait impressions, and how these relative contributions might vary across different traits or contexts, has yet to be established. In a pioneering study, Hönekopp quantified target and perceiver variation for judgments of facial attractiveness, arguing that quantifying this variation is crucial to building complete theory (Hönekopp, 2006). Despite the traditional view that attractiveness is largely a property of the target, and thus more or less universally shared (see Little, Jones, & DeBruine, 2011; Rhodes, 2006 for reviews), Honekopp (2006) found that the variation in judgements of attractiveness was explained as much by the perceiver as by the target, providing new insight into the age-old question of whether beauty is in the eye of the beholder (see also Germine et al., 2015). Remarkably, this approach has so far been limited to attractiveness. Although attractiveness is an important social judgement, perceivers also go beyond impressions of appearance and also readily form impressions of character from targets, such as trustworthiness or dominanance (Oosterhof & Todorov, 2008).

Across a broad array of domains, from personality, social cognition, and impression formation, to visual and auditory social perception, researchers use trait judgments as a common methodological tool. The primary theoretical contribution of the present research is in decomposing these trait impressions, providing evidence to what extent they are in “our minds” vs. “others’ faces”, or in between. To understand how social perception unfolds is to understand what ingredients compose a trait impression, and how they combine. Thus, examining to what extent perceiver- and target-level characteristics contribute to trait impressions will provide important insight into its mechanisms of impression formation, ultimately contributing to a better understanding of the nature of our impressions.

Multilevel Models and Intraclass Correlation Coefficients

One way to decompose the variability in impressions from perceiver and target is to use multilevel models. These statistical models have the advantage over traditional linear regression in that, when repeated observations (e.g., impressions of different targets) are nested within larger clusters (e.g., impressions made by the same perceivers), they can parse what percentage of variance in a dependent variable comes from different levels of the model (Raudenbush & Bryk, 2002). Failing to account for the nested nature of the data at both the perceiver and target level can lead to biased estimates, and effects become an uninterpretable blend of target and perceiver variation (Judd, Westfall, & Kenny, 2012). In the current context, multilevel models provide an elegant statistical avenue to quantify the extent to which an impression stems from the target versus the perceiver. Specifically, with cross-classified multilevel models, we can estimate an intraclass correlation coefficient (ICC) for both the perceiver and for the target on ratings of trait impressions.

These ICCs provide an ideal metric for describing the percentage of variance in a trait rating explained by perceivers and targets. Previous work in person perception has largely relied upon high values of coefficient alpha (Cronbach, 1951) to represent high perceiver agreement in rating targets. Alpha represents an expected correlation between obtained target ratings and a second set of target ratings from an equally large sample of perceivers. However, as discussed elsewhere (Flake, Pek, & Hehman, 2017; Hönekopp, 2006), high alphas are not satisfactory evidence of high perceiver agreement because alpha is strongly influenced by the number of items (here, perceivers). Even weakly correlated ratings of targets will have high alphas provided enough perceivers are included (Cortina, 1993).

In contrast, multilevel models can estimate the variance in a dependent variable that occurs between different clustering variables. Here, these would be multiple ratings made by a single perceiver (i.e., clustered within a single perceiver), and multiple ratings made of a single target (i.e., clustered within a single target). In the same statistical model, we can estimate the variance that is attributable to the perceiver, the variance that is attributable to the target, and (with repeated ratings) the variance attributable to the interaction between targets and perceivers.

The Current Methodological Approach

As described above, with cross-classified multilevel models, we can estimate an intraclass correlation coefficient (ICC) for both the perceiver and for the targets on ratings of trait impressions. The *perceiver-ICC* represents the percentage of variance in ratings that comes from between-perceiver variability (i.e., variability in the characteristics of different perceivers), which might be present due to stable perceiver trait differences or temporary factors (e.g., arousal). The *target-ICC* represents the percentage of variance in ratings that comes from between-target characteristics (i.e., variability in the appearance of targets). The *interaction-ICC* represents the percentage of variance that is due to the unique interplay between targets and perceivers (i.e. personal taste). For example, one perceiver might find people with brown eyes particularly attractive, but not people with blue eyes. Another perceiver might feel the opposite. The attractiveness judgments in this example arise from interactions between perceiver preferences and target characteristics.

Understanding what percentage of variance comes from the perceiver- and target-level is essential to understanding the foundations of different trait impressions. For instance, suppose a perceiver-ICC was .95. This result would indicate that 95% of the variance in a particular trait impression is due to a consistent effect of perceiver-level characteristics, suggesting that people were primarily drawing upon their own mental states to inform their judgments. In contrast, if perceiver-ICCs were only .01, only 1% of the variance in ratings of a trait impression comes from perceiver-level characteristics, suggesting that the appearance of targets was primarily driving the ratings. In this second example, no matter how many perceiver-level variables are included in the model, they will together explain *at most* 1% of the variance in this trait impression. In this hypothetical example, future research would be most usefully directed toward examining visual cues in the faces themselves to explain any effect. Of course, ICCs do not identify *which* perceiver- or target-level variables might best explain a dependent variable. However, they do quantify to what extent variance comes from different levels, and therefore how to develop future theoretical models to best explain that variance.

The Current Research

In sum, in the current work we estimated perceiver- and target-ICCs for different trait impressions to quantify to what extent perceiver- and target-level factors are responsible for final trait impressions. We identify five key questions unanswered by extant models of person perception that have yet to specify the extent to which impressions are driven by perceiver- and target-level characteristics. The first three questions concern how perceiver- and target-level characteristics contribute to distinct trait judgments and dimensions of social judgment. The final two questions concern moderators, or how characteristics of the face or context can moderate perceiver and target contributions to social judgment more generally. Below we outline our specific hypotheses.

Part 1: Distinct Traits and Dimensions of Person Perception

Different Social Perceptions. Because the involvement of perceiver and target characteristics in these different traits has not been quantified and is not considered in most statistical or theoretical models, there is an implicit, functional assumption that perceiver and target characteristics are influencing impressions similarly across different traits. However, it is likely there is substantial variation across different traits, though this has never been examined. Estimating the perceiver- and target-ICCs will reveal which impressions are driven primarily by perceiver-characteristics, which are driven more so by physical target-characteristics, which impressions demonstrate similar structures, and which impressions diverge.

By thus examining different social perceptions by these ICCs, we provide the first test of whether different impressions have different "footprints." That is, are some impressions largely informed by perceiver-level characteristics, whereas others are driven more so by target-level characteristics? We estimate the perceiver- and target-ICCs of 29 trait impressions, chosen because of their common usage and theoretical importance within the person perception literature.

Dimensions Underlying Person Perception. Our next set of analyses aimed to test whether perceiver- and target-ICCs are different across the different dimensions underlying face perception (Oosterhof & Todorov, 2008). Individuals can, of course, be evaluated on a vast number of traits. However, across many different domains, researchers using data-reduction approaches have converged on a smaller set of two or three underlying latent dimensions that explain the majority of the variance in social perceptions (Fiske, Cuddy, Glick, & Xu, 2002; Freedman, Leary, Ossario, & Coffey, 1953; Leary, 1957; Oosterhof & Todorov, 2008; Sutherland et al., 2013). The first dimension is regularly interpreted as whether the target's intentions toward the perceiver are friendly or hostile (Fiske et al., 2002; Oosterhof & Todorov, 2008). The second factor is routinely interpreted as the target's ability to enact those intentions (Oosterhof & Todorov, 2008; Sutherland et al., 2013). These dimensions have been given many different labels across research domains. With respect to face perception, they are commonly referred to as "trustworthiness" and "dominance" respectively (Oosterhof & Todorov, 2008), thus we use these labels for clarity. More recent research incorporating a more variable set of faces further identified an additional factor, “youthful-attractive”, which may have emerged partially due to a broader-aged sample than previous work (Sutherland et al., 2013; Wolffhechel et al., 2014). As the current stimuli were similar in heterogeneity to this more recent work, we included this third dimension in our analyses.

Previous research examining trustworthiness and dominance demonstrated that perceptions of a target’s intentions (trustworthiness) were more variable than perceptions of their dominance (Hehman, Flake, et al., 2015). Further, facial cues underlying these dimensions may differ in salience (e.g., emotional expressions, more salient than other facial cues, may relate most to trustworthiness perceptions; Hansen & Hansen, 1988). Thus, because the cues to each dimension differ in variability and salience, we expected a larger contribution of target variation to the dimension of trustworthiness, compared to dominance.

Our expectations for youthful-attractiveness were less clear. Recent work has revealed there is a surprising amount of variability across individual perceivers in what is considered attractive (Germine et al., 2015; Hönekopp, 2006). This individual variability might be reflected in a particularly large *perceiver* variance in impressions of the youthful/attractive dimension. Yet, the third youthful-attractiveness dimension underlying person perception also depends on cues to age (Sutherland et al., 2013). Because age is conveyed by many cues in the face and is fairly veridical, the *target* variance for the youthful/attractive dimension might instead be particularly high. We compared perceiver- and target-ICCs for these three different dimensions.

Perceiver × Target Interactions. In the majority of data comprising the current research, participants rated each target only once. This data structure did not allow for the estimation of the random effect associated with the perceiver by target interaction (described more fully below). Many theories in social cognition, however, depend on impressions being jointly influenced by both perceiver and target characteristics, and examining the extent to which impressions are driven by blends of both perceiver and target characteristics would reveal a host of implications for person perception models. For instance, to what extent do perceiver characteristics (e.g., sexism) uniquely influence ratings of some targets (e.g., female) but not others (e.g., male)? We predicted that a substantial percentage of variance in impressions would stem from perceiver × target interactions, suggesting that impressions are differentially formed across perceiver and target pairs.

Examining interactions between perceiver and target characteristics required a unique data structure not present in the majority of data analyzed in the present research (or indeed, in the majority of the field). We therefore collected data in which participants rated each face twice, such that the variance of the perceiver × target interaction could be estimated (details below). This approach allowed us in Analysis 3 to quantify the extent to which impressions were unique blends of perceiver and target factors simultaneously on each dimension. Our results suggest that domains of social judgment are likely more complex than previously realized.

Part 2: Moderators of Perceiver and Target Contributions to Judgments

While our first three research questions above examine variability across different traits and dimensions, our latter two examine how this variability can be moderated across different contexts. Specifically, how characteristics of the face or context can change perceiver and target contributions to social judgment more generally. We propose that even across the diversity of traits on which perceivers form impressions of others, contextual and ambient factors influence the breakdown of perceiver and target contributions to those ratings.

Extremity of Emotional Expression. For example, we predict that as emotional displays on faces become more extreme, there is less room for interpretation of even non-emotion judgments (i.e., decreasing perceiver variance). More emotionally neutral displays may invite more perceiver variance in impressions, compared to faces with greater emotional extremity. Such a test is theoretically interesting with respect to the emotion display literature, while also providing a validation of our overarching hypothesis that perceivers contribute more to more ambiguous evaluations. Accordingly, we hypothesized that perceiver-ICCs would be greater when emotional expressions of faces were ostensibly neutral, as compared to faces with more extreme displays of emotion.

Real vs. Computer-Generated Faces. Finally, we ask another important question for research: does using computer-generated stimuli change the perceptual process? A great deal of person perception research uses software to create computer-generated faces for research (e.g., FaceGen; Blanz & Vetter, 1999), as it offers fine-grained experimental control. An obvious concern when using these faces is whether the conclusions generalize to real faces (Crookes et al., 2015). As faces become more standardized (whether via using controlled photographs or even by computer generation), attention might be more focused on certain facial features (i.e., increasing target variance). Our final set of analyses test whether perceiver- and target-level characteristics contribute equally across impressions of both real and computer-generated faces.

Finding moderators of perceiver- and target- contributions to trait judgments, broadly construed, would suggest that the sources of variance in domain-general social judgment can be swayed by contextual and ambient factors.

Summary of Current Approach

In summary, in five different analyses we examined differences in how perceiver variability and target characteristics contribute to impressions across 1) an array of theoretically important judgments, 2) the core dimensions of person perception, 3) perceiver × target interactions, 4) extreme vs. neutral emotional expressions, and 5) real vs. computer-generated faces. To do so, we partitioned a large database of ratings as a function of the question. We report each of the five analyses as if each were a separate study in a multi-study paper, detailing the participants, stimuli, ratings included, and results.

Methods

Analytical Approach

Across all analyses, we ran a series of multilevel models to calculate the intraclass correlation coefficients (ICCs). In these models the trait or dimension being evaluated (e.g., friendliness) acts as the single dependent variable. The variance in ratings of that trait is decomposed into distinct parts: that attributable to the target, the perceiver, the perceiver × target interaction (when we have repeated measures within perceiver) and what is left over (i.e., the residual or error variance). This model is called a null or intercept-only model because it does not include any independent variables or covariates. Our models are also cross-classified, in that ratings were nested within both participants and targets (Judd et al., 2012; Raudenbush & Bryk, 2002). Accordingly, an ICC for both the perceiver and target can be calculated.

Formally, the multilevel model can be represented with two equations, one for the first level of the model and the other for the second:

$$\text{Level 1: } Y_{i(j_1 j_2)} = \pi_{0(j_1 j_2)} + e_{i(j_1 j_2)}$$

$$\text{Level 2: } \pi_{0(j_1 j_2)} = \theta_{000} + b_{0 j_1 0} + c_{00 j_2} + d_{0(j_1 j_2)}$$

In the first level of the model, $Y_{i(j1j2)}$ is the dependent variable, which for our purposes is a rating of a trait i (e.g., friendliness) of target j_1 by perceiver j_2. The intercept in this model, $\pi_{0(j1j2)}$, is the expected value of the rating from target j_1 by perceiver j_2. The error term, $e_{i(j1j2)}$, has associated variance, σ^2. In the second level of the model, the intercept is modeled as an outcome that varies across targets and perceivers, allowing the decomposition of the total variance into that attributable to the perceiver and target. θ_{000} represents the grand mean, or the average rating across all targets and perceivers. From that grand mean, b_{0j10} represents the residual, or the difference between this grand mean and the rating of target j_1 averaged across all perceivers; these residuals have variance τ_{b00}. c_{00j2} represents the residual of perceiver j_2 averaged across all targets, which has variance τ_{c00}. The final random effect, $d_{0(j1j2)}$ represents the interaction, or the variance that comes from the unique combinations of targets and perceivers. The variance of the interaction term is usually fixed to zero, because it cannot be disentangled from the level 1 error variance without sufficient cell sample size (Beretvas, 2008; Raudenbush & Bryk, 2002). In the context of the current study, estimation of the interaction variance is only possible if a perceiver rates the same targets at least twice (i.e., repeated measures within a perceiver and target). In Analysis 3, we collected data in order to specifically estimate this interaction component.

From these estimates the perceiver- and target-ICCs can be calculated (see Supplementary Materials for example R code and instructions for calculation). For example, the ICC for the target is calculated as a proportion of the total variance that can be attributed to the target:

$$ICC_{\text{target}} = \frac{\tau_{b00}}{\tau_{b00} + \tau_{c00} + \sigma^2}$$

Analyses were conducted in R (lme4: Bates, Mächler, Bolker, & Walker, 2015). Though our goals were largely descriptive, we used two-tailed *z*-score tests for population proportions to test whether ICCs were significantly different from one another.

Source of the Data

A large dataset was necessary such that precise and generalizable estimates of perceiver- and target-ICCs could be obtained. To this end, all data collected by the first author consisting of social perception ratings of facial stimuli were included and aggregated. Across all ratings of social perceptions, targets appeared in random order, and were rated from 1-“Not at all” to 7-“Very much” Likert scales on different traits (e.g., “How friendly is this person?). Participants rated targets on only one trait to avoid crossover effects (Rhodes, 2006). These criteria resulted in 698,829 ratings of trait impressions (e.g., friendly) across 6,593 participants and 3,353 stimuli. Data were collated from participants in a lab along with those recruited on Amazon’s Mechanical Turk between 2011 and 2016 (M_{age}=35.51, *SD*=12.28, 59% female, 77.2% White when race reported[1]). Participant ratings of trait impressions included: aggressive (*n* = 14,569), angry (*n* = 857), assertive (*n* = 15,279), attractive (*n* = 121,960), caring (*n* = 2,740), competent (*n* = 64,559), creative (*n* = 2,020), dominant (*n* = 77,300), feminine (*n* = 9,976), friendly (*n* = 80,903), gender-typical (*n* = 4,240), happy (*n* = 857), healthy (*n* = 2,800), intelligent (*n* = 63,648), likeable (*n* = 11,214), mean (*n* = 2,020), physically powerful (*n* = 885), race-typical (*n* = 3,901), racist (*n* = 6,884), smart (*n* = 2,847), socially powerful (*n* = 1,416), physically strong (*n* = 79,379), trustworthy (*n* = 60,383), warm (*n* = 42,158), wise (*n* = 10,133), and youthful (*n* = 15,901). Data were collected across 39 different studies, in projects both published and unpublished.

Stimuli

An important factor to consider when estimating the percentages of variance from the perceiver- and target-level is the overall variance in the set of stimuli on each trait. For instance, consider participants rating the attractiveness of a group of fashion models vs. participants rating a wider, more representative, sample of targets. Previous research has illustrated that low variance in the attractiveness (in this case) of the targets yields artificially higher perceiver-ICCs for this impression (Hönekopp, 2006). Thus, to provide generalizable estimates of perceiver and target-ICCs, we considered it essential that the sample was large and heterogeneous in its representation of diverse traits. Others have made similar arguments for data driven approaches using heterogeneous naturalistic stimuli (Burton, Kramer, Ritchie, & Jenkins, 2015; Jenkins, White, Van Montfort, & Burton, 2011; Sutherland et al., 2013). The data used in the current work was ideal for this purpose, given that it was curated from a wide variety of sources (e.g., politicians, undergraduate volunteers, baseball players, computer-generated models, mugshots, Facebook profiles, CEOs, Playboy playmates, academic databases, fraternities, etc.) to test different hypotheses. Examples of the different stimuli are provided in Figure 1.

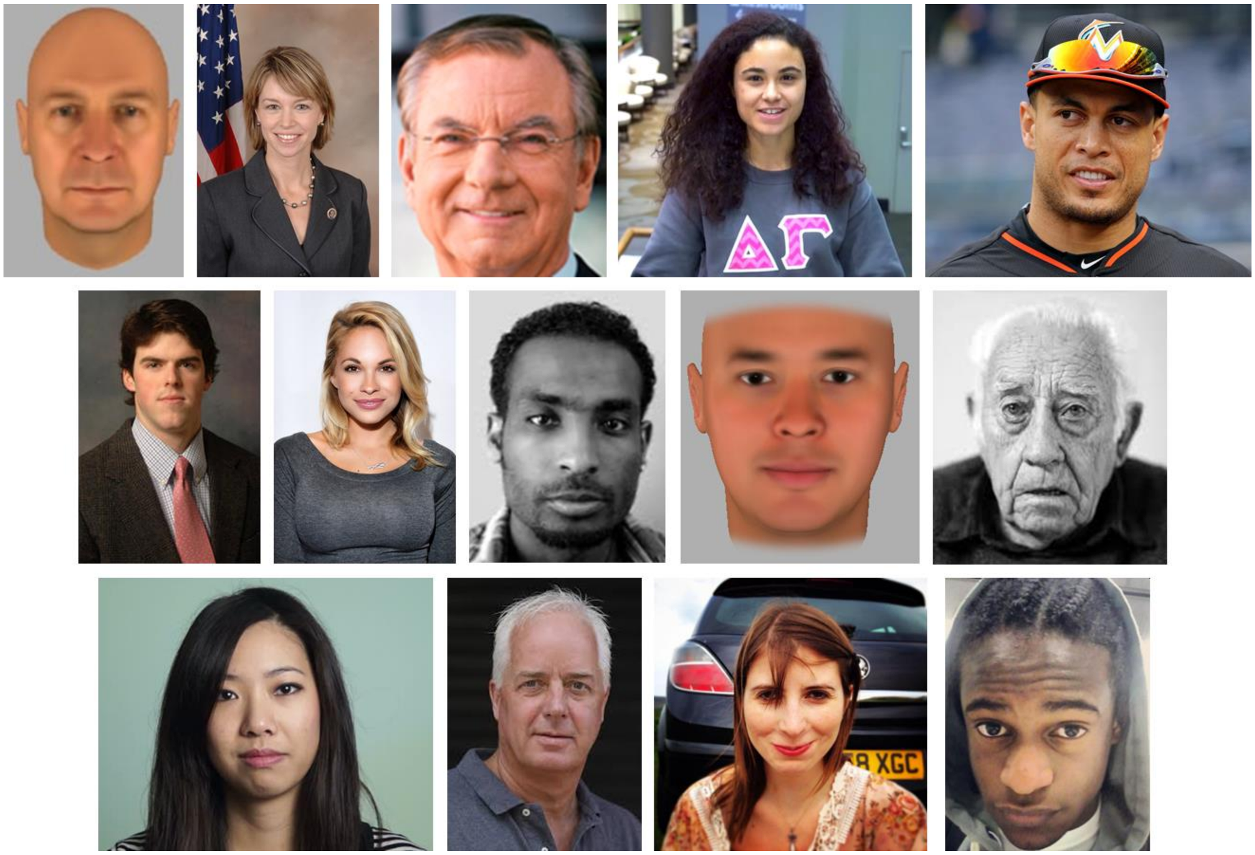

Figure 1. Example stimuli.

The faces represented a wide range of facial variation as is typically studied in psychological experiments as well as images as encountered in real life or when browsing the Internet, offering an ideal starting place for our central question of how important perceiver and target variation are in facial impressions. Thus, our data had the heterogeneity necessary to allow our estimates to generalize beyond the sample.

Part 1: Distinct Traits and Dimensions of Person Perception

In the first part of the paper we examine perceiver- and target-ICCs for a variety of traits. Analysis 1 reveals that the origins of variance in traits are diverse. Perceiver and target characteristics do not influence impressions similarly across different traits, as implicitly assumed by prior work. Next, in Analysis 2 we examine how perceiver- and target-ICCs differ across dimensions of person perception, said to commonly underlie the diverse set of traits examined in Analysis 1. These dimensions show unique patterns of perceiver- and target-ICCs, providing insight into their substrates. Finally, Analysis 3 unpacks impressions with a unique dataset that allowed us to parse perceiver × target effects in judging core person perception dimensions.

Analysis 1: Different Social Perceptions

Results

Figure 2 displays the surprising variability in the extent to which perceiver and target characteristics contribute to impressions of different traits. Bootstrapped correlations indicated that perceiver- and target-ICCs were negatively correlated with one another ($r = -.686$, $p = .0002$, 95% CI [-.833, -.396]), but were unrelated to the number of observations, participants, or stimuli involved in each analysis (all *p*s > .1).

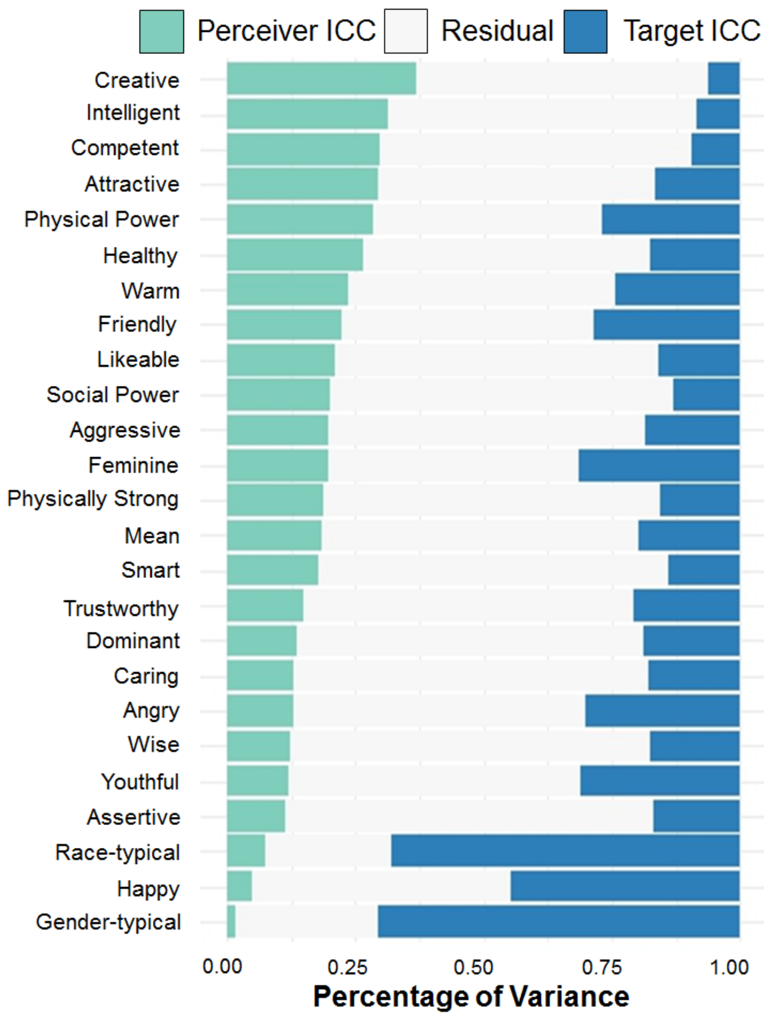

Figure 2. Relative contributions of between perceiver (Perceiver-ICC), between target (Target-ICC), and within perceiver and target variance (Residual) to all trait impressions in Analysis 1.

Discussion

The pattern of results from Analysis 1 offers a host of theoretical implications for future research. Importantly, these results make clear that perceiver and target characteristics do not influence impressions similarly across different traits, as implicitly assumed by prior work. Impressions with larger target-ICCs are being driven to a larger extent by target-level characteristics, suggesting that certain facial features are responsible for impressions, with little room for perceiver interpretation. For impressions such as race-typical and gender-typical, perceivers appear to readily agree whether a face is typical for that social category, and thus which facial features covary with social categories. The higher target-ICC for youthful similarly indicates that perceivers agree on features determining this judgment (likely whether signals of age are present or not) and again with judgments of happiness and anger (likely whether faces resemble happy or angry expressions, respectively).

What the above impressions share is they are all appearance-based appraisals, and yet some inference-based trait impressions demonstrate similar patterns, yielding insight into how these inferences into character might share similar origins. For instance, friendliness has the highest target-ICC of these inferences, suggesting that perceivers agree on which facial features convey friendliness, and that people are treating this judgment not unlike judgments of happiness or anger. In other words, people are likely using facial features that resemble emotional expressions for these judgments. In contrast, ratings of creativity have the lowest target-ICC, suggesting that raters show very little agreement about which facial features convey creativity.

Conversely, the magnitude of the perceiver-ICC reveals unique groupings of these trait ratings, revealing to what extent perceiver-level factors color impressions. For example, creativity has the highest perceiver-ICC, suggesting that individuals draw upon their personal understandings of creativity to make such judgments, with some perceivers consistently rating all faces higher than other perceivers. Judgments of intelligence and competence similarly seem to leave room for perceiver interpretation. Yet for other, content-similar traits (e.g., wise), perceiver-factors play a smaller role.

Thus the present pattern of trait ratings provides insight into the extent to which traits are expressed reliably via facial cues, or in contrast, those which rely upon perceiver inferences. This mapping provides a different way to think about impression formation. That is, rather than construing person perception along a content-based space of broad domains of judgment (e.g., competence and warmth, or trustworthiness and dominance)—which features prominently in social cognition—we could instead think of trait judgments in an alternative space: how much perceivers bring to bear in forming judgments, or how much the target displays features consistently elicit a judgment. Further, these two approaches (i.e., dimensions of social judgment, and sources of social judgment) can be integrated in theoretically meaningful ways, to which we turn next.

Analysis 2: Dimensions Underlying Person Perception

Results

Core person perception dimensions. Our second aim was to compare perceiver and target characteristics in their contribution to the major dimensions underlying person perception. We first had to create the underlying dimensions. Because ratings were from numerous different participants and stimuli across different studies, conducting comprehensive factor analyses on these data to derive dimensions was not possible. Fortunately, several large-scale factor analyses of trait impressions from faces have been conducted, and we created our dimensions based on these studies and the large amount of subsequent research supporting these conclusions.

Initial groundbreaking work with controlled stimuli found two dimensions of social judgment underlie impressions of faces (Oosterhof & Todorov, 2008), one representing trustworthiness and the other dominance. Subsequent research with a broader stimuli set, including targets with more variable ages, replicated this work but additionally found a novel dimension representing youthful/attractive (Sutherland et al., 2013; Wolffhechel et al., 2014). Because our sample was highly heterogeneous and included older aged targets, we also included the youthful/attractiveness dimension. Thus, we used this previous research to map different traits to different underlying dimensions.

Ratings for the 20 traits ultimately included in calculating the ICCs for each dimension were: trustworthiness (aggressive, caring, creative, friendly, likeable, trustworthy, warm, wise), dominance (assertive, competent, dominant, intelligent, mean, physically powerful, physically strong, smart, socially powerful), youthful/attractive (attractive, healthy, youthful). Across the three dimensions, 664,321 ratings were made across 6,985 participants and 3,069 stimuli (see the Supplementary Materials for a correlation matrix representing relationships between all traits).

Person perception dimensions analysis. Averaging across all three dimensions, perceiver variability contributed 22.8% of the variance whereas target characteristics contributed 17.7%.

Perceiver-level ICCs. The dimension with the greatest amount of variance explained by perceiver-level characteristics was youthful/attractive (*perceiver-ICC* = .279), which was significantly greater than both trustworthiness (*perceiver-ICC* = .195), $z = 5.95$, $p<.0001$, and dominance (*perceiver-ICC* = .210), $z = 5.32$, $p<.0001$. Trustworthiness and dominance did not differ, $z = 1.29$, $p = .1971$ (Figure 3).

Target-level ICCs. The greatest amount of variance explained by target-level characteristics was trustworthiness (*target-ICC* = .234), followed by youthful/attractive (*target-ICC* = .165), followed by dominance (*target-ICC* = .131). Each target-ICC was significantly different from all others, all *z*s > 3.48, all *p*s < .0005.

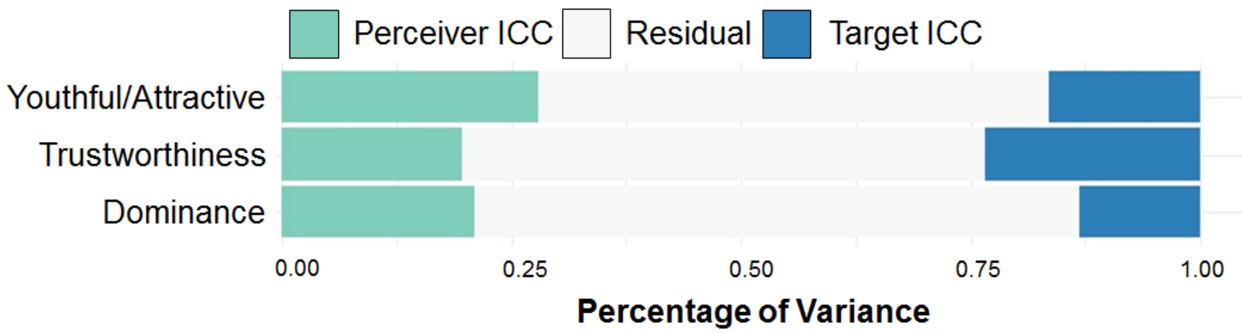

Figure 3. Relative contributions of between perceiver, between target, and within perceiver and target variance to impressions across the dimensions underlying person perception in Analysis 3.

Discussion

Previous research has posited that trustworthiness, dominance, and youthful/attractiveness are distinct dimensions in person perception, and that we find a distinct footprint of perceiver- and target-level contributions to impressions for each of these different dimensions supports this conclusion. Importantly, these results suggest that the causal process of forming impressions along each of these dimensions is relatively unique.

In particular, characteristics of the *target* were especially important for trait impressions of trustworthiness (23.4%). This result indicates that target-level variation has a greater impact on ratings along the trustworthiness dimension than dominance or youthful-attractiveness dimensions. One possible explanation for this result is that the facial cues that inform ratings of trustworthiness might be especially salient. Previous research has demonstrated that perceptions of trustworthiness largely rise from emotional expressions (Adams et al., 2012; Oosterhof & Todorov, 2009; Said et al., 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003; Zebrowitz et al., 2010), which can be especially salient when perceiving faces (Hansen & Hansen, 1988). In contrast, perceptions of dominance appear to be driven more by facial morphology such as a wider face (relative to its height) or larger brow (Carré et al., 2010; Hehman, Leitner, & Gaertner, 2013) though see Sutherland et al., 2016; Zebrowitz et al., 2010 for evidence that expressions can contribute to evaluations of dominance), and perceptions of youthful/attractiveness by changes in facial morphology or texture with aging (Hehman, Leitner, & Freeman, 2014; Sutherland et al., 2013). If temporary emotional expressions are indeed more salient than stable facial morphological cues, this pattern would explain the current results.

Results further indicate high variability in overall trait impressions of youthful/attractive across perceivers. This is broadly consistent with research demonstrating that there is a great deal of idiosyncratic variability in perceptions of attractiveness across individuals (Germine et al., 2015; Hönekopp, 2006). However, it is hard to directly compare estimates with these previous findings, which mainly focused on perceiver variation as the *interaction* between perceivers and targets.

Indeed, here, potential interactions between perceivers and targets is inextricably entangled with the level 1 residual variance. This residual varies in magnitude across the three dimensions, largest for dominance and smaller for the other two. This result indicates that ratings along the dominance dimension potentially have larger perceiver × target interplay. However, because of our data structure (i.e., one rating per participant per target), we cannot separate interactions from the level 1 residual, and thus it is difficult to interpret differences in the residual across dimensions. Accordingly, our next analyses turned to these interactions.

Analysis 3: Describing Variability from Target by Perceiver Interactions

For all trait ratings above, participants rated each target one time. This data structure did not allow for the estimation of the random effect associated with the perceiver by target interaction because the rating was not repeated within participant for a single target. However, with multiple ratings of the same target by the same participant, the variance associated with the interaction can be parsed from the residual variance. Here we present a model where the error term, σ^2, represents the variance in the reliability of the two ratings across people. In the second level of the model, we now additionally estimate the random effect, $d_{0(j1j2)}$, which represents the interaction, or the variance that comes from the unique combinations of targets and perceivers, after taking into account the perceiver and target main effects.

Methods

New participants (*n* = 211) recruited on Mechanical Turk rated 50 White male and female faces from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) from 1-“Not at all” to 7-“Very much” on 6 different traits. For this analysis, we selected the two traits loading most strongly on each dimension from Analysis 2: friendliness, trustworthiness, physical strength, dominance, youthfulness, and attractiveness. Target faces appeared in random order, and participants rated faces on only one trait, providing two ratings for each face (full set of 50 faces for a total of 100 trials). This approach resulted in 20,133 ratings from which we calculated a perceiver-ICC, target-ICC, perceiver × target interaction-ICC and the level 1 residual. Traits loading on the same dimension were combined.

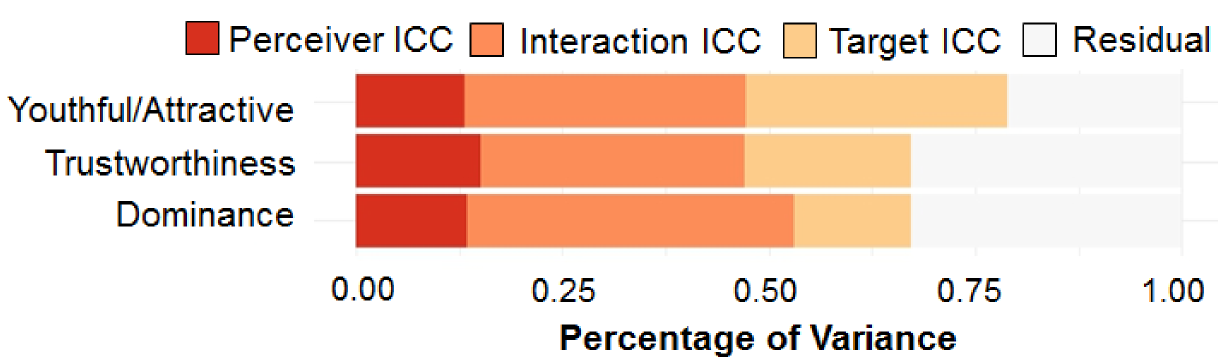

Figure 4. Relative contributions of between perceiver (Perceiver-ICC), between target (Target-ICC), between perceiver × target combinations (Interaction-ICC), and the residual to all dimensions in Analysis 3.

Results and Discussion

The new Interaction ICC can be interpreted as the percentage of variance that is attributable to the unique combination of perceiver and target characteristics, beyond the main effect variance of targets or perceivers. Crucially, we can see that for each dimension, a substantial percentage of the variance in ratings is a result of meaningful interactions between perceiver and target characteristics. Though the percentage of variance attributable to the interaction ranges from 32.1% (Trustworthiness) to 39.5% (Dominance), in each case it is substantial. This pattern clearly supports our overall point that perceivers also actively interpret social targets and that future research needs to consider this variation.

We can also now more directly compare our youthful-attractiveness estimate to previous studies examining attractiveness (Germine et al., 2015; Hönekopp, 2006). Our findings conceptually replicate this previous research. We similarly found that variation in the youthful-attractiveness dimension is equally due to interactions between the perceiver and the face (*Interaction-ICC:* 34.1%), which could be called "personal taste", relative to the face alone (*Target-ICC:* 31.6%), representing consensually agreed-upon elements of attractiveness. The remaining perceiver variance represents a main effect of perceivers (e.g., some perceivers consistently judging faces higher on attractiveness).

Critically, we also extend this previous work by showing that the other two dimensions, representing inferences of character rather than appearance, are even more influenced by this interaction between the perceiver and target, indicating that there is more to learn about the nature of these judgments. In particular, different perceivers may use different cues to form these impressions, especially for dominance (returned to in the General Discussion).

The results displayed in Figure 4 can be compared with that of Figure 3 to examine how perceiver- and target-ICCs differ when perceiver × target interaction is disentangled from residual variation. We find key similarities and interesting differences across analyses. First, dominance is clearly still the least target-led dimension, as before. Moreover, the new data now further reveal that dominance shows the largest perceiver by target interaction variance, meaning that different perceivers appear to be judging dominance from faces differently (as well as differing in their overall dominance impressions). However, unlike in Analysis 2, youthful-attractiveness is now the most target-led (and least perceiver-led) dimension, with trustworthiness falling in between.

Some differences in the target and perceiver-ICC calculations should be expected. These analyses are based on a smaller and less heterogeneous sample relative to the rest of the paper. ICCs (especially target-ICCs) are impacted by the overall amount of stimuli variance (Hönekopp, 2006), and so different stimuli variance may be involved in any ICC differences between Analysis 2 and 3. The face stimuli here were also emotionally neutral, unlike in Analysis 2, and removing emotional expression would contribute to a lower target ICC for trustworthiness, given the importance of this cue for judging this dimension. Moreover, with repeated ratings comes distinct psychological phenomenon due to other known tendencies such as mere exposure, familiarity, halo effects, and perceptual recalibration (Lorenzo, Biesanz, & Human, 2010; Nisbett & Wilson, 1977; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Zajonc, 1968). These phenomena may also change the extent to which perceiver and target characteristics, and their interplay, drive specific ratings. We hope our analysis inspires future studies to systematically test these effects.

Finally, we note that the residual values from these analyses are of greater utility here, as they now form a measure of reliability. Specifically, they represent variance across people in the discrepancy between their two ratings of the same target, with lower variance indicating greater consistency. For example, these results indicate that people are more consistent in their ratings of the youthfulness/attractiveness dimension (21.1%) across repeated ratings, relative to the other two dimensions. It is an interesting question for future research as to the optimal number of repeated ratings. Increasing repetitions of ratings allow for more stable estimates of reliability (Nunnally, 1978). However, researchers interested in "first impressions" may face a limit on the repetitions that are possible, given that repeated exposures may change the phenomenon of interest in qualitatively meaningful ways. Again, our analysis opens up these questions as interesting new research avenues.

Supplementary Analysis: Ease of Rating

Why is there such a great deal of variation in perceiver- and target-ICCs across ratings of different traits? The many theoretically important reasons for perceiver variation described in the introduction are too vast to systematically test here. However, one untested possibility is that different patterns of variance stem from participants finding some traits more difficult to evaluate than others, and if so, participants may themselves be consciously aware of this difficulty. Alternatively, participants might be unaware of the extent to which ratings of different traits are idiosyncratic to the perceiver and target.

Methods. We therefore asked new participants (n = 132) recruited on Mechanical Turk to rate the perceived ease of evaluating faces on different traits. On a 1-"Not at all easy" to 7-"Very easy" scale, participants responded to the question, "If you were just looking at someone's face, how easy would it be to tell how [trait] they are?", for all the traits included in the present research. Trait order was randomized by participant. Ratings were averaged for each trait such that trait operated as the unit of analysis. We then correlated these averaged easiness ratings of each trait from this new sample with the ICCs of each trait, estimated from the large main sample in Analysis 1.

Results and Discussion. Bootstrapped correlations indicated that while rated easiness of impressions was uncorrelated with perceiver-ICC (r = -.298, p = .1483, 95% CI [-.680, .229]), it was positively correlated with target-ICC (r = .616, p = .0010, 95% CI [.343, .877]). Thus, as participant meta-perceptions regarding the ease of rating different traits increased, so too did the extent to which target-level characteristics drove the impressions. Plotting this correlation (Figure 5) reveals some interesting discrepancies in the mismatch between rated ease of impressions and target-ICCs. In Figure 5, above the dotted line (representing a correlation of r = 1.00), are trait impressions that are apparently driven *more* by target-characteristics than participants believed (e.g., race-typical, gender-typical). Below the dotted line are trait impressions that are apparently driven *less* by target-characteristics than participants believed (e.g., attractive, youthful).

The positive correlation between ease of rating a trait and target-ICC suggests that participants are generally aware of how difficult it may be to rate targets on more ambiguous traits. Yet some notable exceptions (e.g., race-typical, attractive) highlight impressions for which participants are incorrect in the extent to which impressions are target driven.

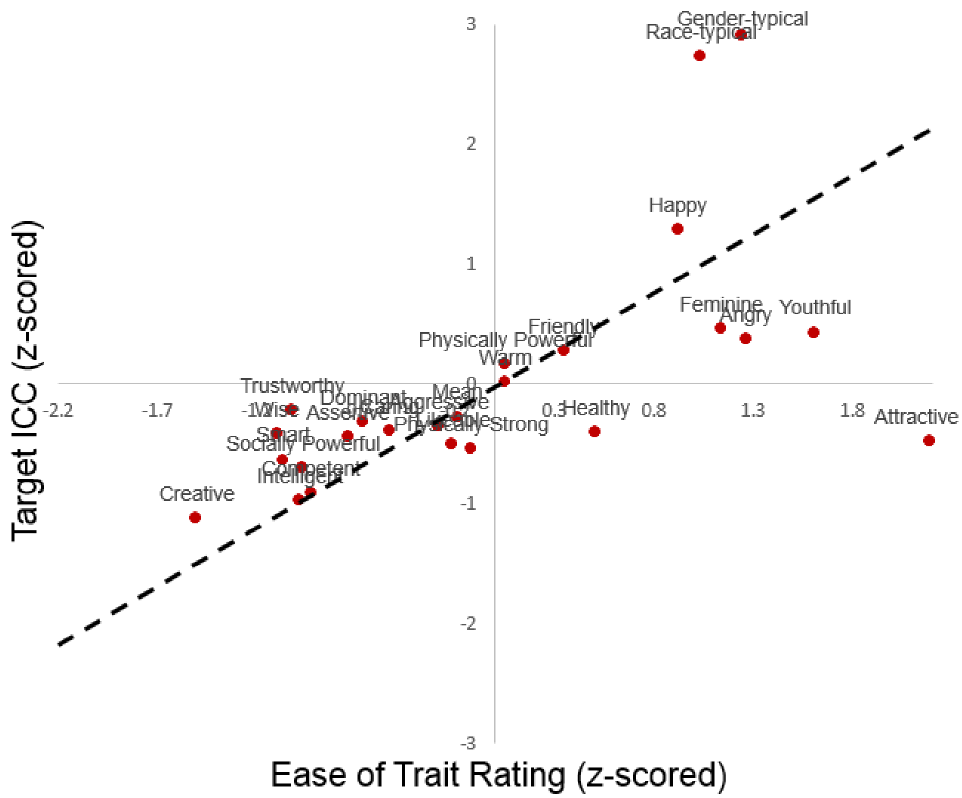

Figure 5. Scatterplot of *z*-scored rated ease of ratings and *z*-scored target-ICC for each trait. The dotted line represents a correlation of $r = 1$.

Discussion

Examining how perceiver variability and target characteristics influence different impressions (Figures 2-4) reveals surprising variability in the "footprints" of different traits and dimensions. Perceiver variability contributed from 1.8% to 36.8% of the variance in different impressions. Target characteristics contributed from 6.2% to 70.6%. That perceiver- and target-ICCs were negatively correlated indicates that, at least in making ratings of others' faces, there is some trade-off between characteristics of the target and perceiver characteristics in forming impressions. We also conceptually replicate the finding that impressions of (here, youthful-) attractiveness are driven as much by the unique interplay between perceivers and targets as variation in the targets themselves (Germine et al., 2015; Hönekopp, 2006). Importantly, we also find that the other two dimensions of social perception, trustworthiness and dominance, are even more driven by this interplay between perceivers and faces. In general, across analyses, we find that traits or dimensions that require more inference (e.g. creativity impressions, the dominance dimension) are less target- and more perceiver-driven. To our knowledge, these differences across impressions and dimensions have never been documented, and have important methodological and theoretical implications (see General Discussion).

Importantly, statistical models that do not account for this variation are ignoring important mechanisms of social perception, as well as making an implicit yet functional assumption that perceiver- and target-characteristics contribute equally to different trait impressions, which the present results reveal is incorrect. Previous models may have conflated these unique contributions due to inflexible models, but the present research demonstrates that statistical models now exist that can appropriately model the complexity inherent in impression formation.

Part 2: Moderators of Perceiver and Target Contributions to Judgments

Whereas Part 1 examined variance in specific traits or domains of judgment, Part 2 used this same analytic technique to examine how the sources of variance in social judgment might vary across different sets of face images. For instance, Analysis 4 tests whether perceiver and target characteristics play larger or smaller roles depending on the emotional extremity of a face. Moreover, perhaps even the features of the image itself may influence judgments. For instance, Analysis 5 examines whether naturalistic images, that is those taken in highly unstandardized settings, may allow for greater perceiver interpretation than standardized images (e.g., photo databases, computer-generated faces) that may focus perceivers more on specific target facial features, limiting perceiver contributions and increasing target contributions to impressions.

Though perceiver × target effects may be present in these analyses, due to the present data structure we were not able to explore this possibility. Therefore, like Analysis 1 and 2, in Analysis 4 and 5, potential perceiver × targets interactions are included in the level 1 residual variance.

Analysis 4: Extremity of Emotional Expression

Methods

Stimuli were computer-generated and manipulated to appear displaying emotion on a 5-point continuum from subtly angry expressions to neutral to subtly happy expressions (see Supplementary Figure 1 for stimuli example). These five levels (e.g., -2, -1, 0, +1, +2) were recoded as three levels of emotional expression intensity: high, medium, and low (i.e., using the absolute value). Specifically, the happiest and angriest faces were recoded to high, the moderately happy and angry faces recoded to medium, and the neutral faces recoded to low emotional intensity.

Because the faces used in these particular analyses were all computer-generated, they were controlled to display equally intense emotional expressions across different target identities. Therefore, differences in target-ICC were not expected for the current data as there was no variance in emotional expression *within* each category of emotional extremity (i.e., high, medium, low). This issue is idiosyncratic to the current data, however: other samples may fruitfully explore target-level variation.

The faces included in this analysis were rated on dominance, friendliness, physical strength, trustworthiness, and warmth. We compared ICCs across these different levels. These analyses included 114,919 ratings from 1,374 participants across 772 stimuli.

Results

Emotional intensity analysis. We predicted that faces presented with more extreme emotional expressions would leave less room for perceiver interpretation in impressions, and thus that perceiver-ICC would be lowest for high emotion faces, and highest for low emotion faces.

Perceiver-level ICCs. As predicted, the percentage of variance in ratings from perceiver variability was greater for low emotion faces (*perceiver-ICC* = .262) than for high emotion faces (*perceiver-ICC* = .171), $z = 3.45$, $p = .0005$. Medium emotion faces (*perceiver-ICC* = .212) were not significantly different from high, $z = 1.53$, $p = .1260$, and marginally different from low emotion faces, $z = 1.73$, $p = .0836$ (Figure 5).

Target-level ICCs. As expected, given the controlled face stimuli, the percentage of variance in impressions from target-level variation was not significantly different across any level of emotion, all *zs* < .22, all *ps* > .83.

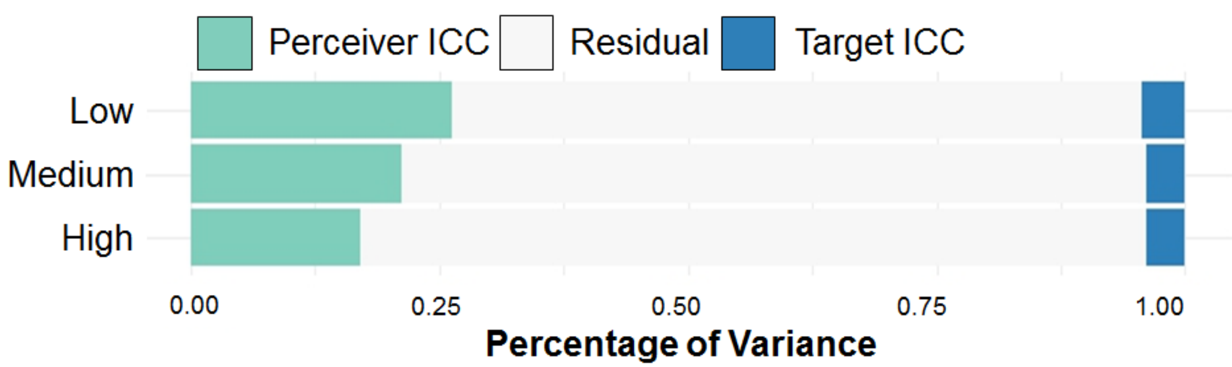

Figure 6. Relative contributions of between perceiver, between target, and within perceiver and target variance to impressions across stimuli varying in the extremity of emotional expression in Analysis 4.

Note. Differences in target-ICC were not expected, due to no variance in emotional expression *within* each category of emotional extremity.

Discussion

Perceiver variability played a greater role in driving impressions of targets with less emotional intensity, as anticipated. Though the range of facial emotion in these stimuli was subtle (Supplementary Figure 1), based on these results we would expect that as the intensity of the emotional expression increased, the variability attributable to the perceiver in the ratings of these faces would decrease further.

Importantly, the ratings examined in this analysis were not judgments of emotion, but rather inferences of dominance, friendliness, physical strength, trustworthiness, and warmth. Yet, the more emotionally neutral a face, the more perceiver variance contributed to these ratings. Thus, people vary more widely in forming social inferences from ostensibly expressionless faces relative to faces with more obvious displays of anger and happiness.

Analysis 5: Real vs. Computer-Generated Faces

With increasing use of computer-generated faces in social cognition research, an obvious concern is the external validity of conclusions drawn from such stimuli relative to real faces. In the current data, 22.4% of faces rated (156,361 ratings) were generated using computer software. Accordingly, we could examine whether perceiver and target characteristics vary across these stimuli types. This analysis included 698,829 ratings across 6,595 participants and 3,359 stimuli.

Results

All photos.

Perceiver level ICCs. The percentage of variance in impressions due to the perceiver was greater for real (*perceiver-ICC* = .237) than computer-generated faces (*perceiver-ICC* = .165), $z = 4.76$, $p<.0001$.

Target-level ICCs. The percentage of variance in impressions from target-level variation was equivalent for real (*target-ICC* = .078) and computer-generated faces (*target-ICC* = .087), $z = .90$, $p = .3681$ (Figure 6).

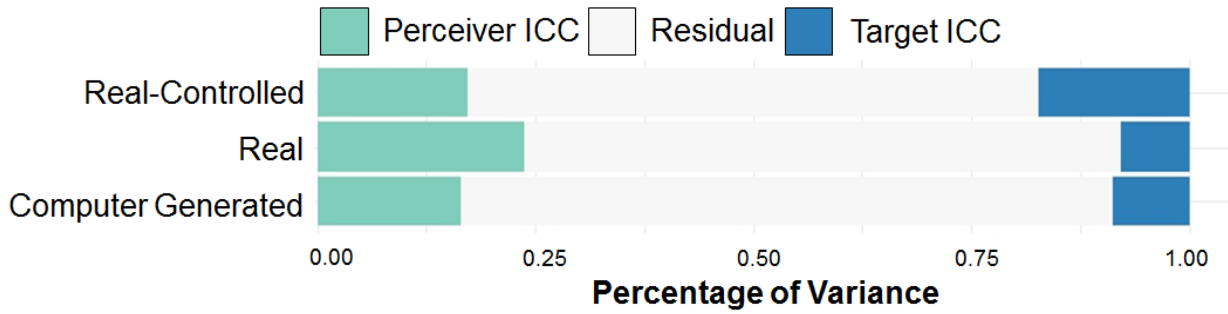

Figure 7. Relative contributions of between perceiver (Perceiver-ICC), between target (Target-ICC), and within perceiver and target variance (Residual) to impressions across computer-generated, real, and real faces from controlled stimuli databases in Analysis 5.

Standardized photos. The computer-generated faces were standardized for expression and pose, front-facing, and displayed with gray backgrounds. In contrast, some of the real faces in the previous analysis highly varied on a number of different potential cues to impressions (e.g. emotional expression, viewpoint, coloring, environment, etc.), because they came from the internet (see Jenkins, White, Van Montfort, & Burton, 2011 for advantages of using naturalistic images). Therefore, perceiver-level characteristics may have a greater opportunity to influence impressions from these naturalistic real faces, due to the increased number of potential cues present. Thus to more fairly compare real and computer-generated faces, we conducted another analysis including only ratings of faces from established face databases that presented the stimuli in controlled and standardized environments. These databases included the Chicago Face Database, the Center for Vital Longevity database, Eberhardt's face database, and the Karolinska Institute database (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Lundqvist, Flykt, & Öhman, 1998; Ma et al., 2015; Minear & Park, 2004). This analysis included 231,858 ratings across 1,914 participants and 1,163 stimuli.

Perceiver-level ICCs. When comparing perceiver variability for impressions of controlled-real faces (*perceiver-ICC* = .173) with that of computer-generated faces (*perceiver-ICC* = .165), there was no significant difference, *z* = .45, *p* = .6527.

Target-level ICCs. With this comparison, the percentage of variance in impressions from target-level variation was greater for controlled-real faces (*target-ICC* = .173) than computer-generated faces (*target-ICC* = .087), *z* = 2.97, *p* = .0030.

Discussion

When comparing a broad range of real faces (both controlled and naturalistic) with computer generated faces, perceiver variability initially appeared to play a larger role in impressions of real faces. When a more comparable set of standardized real face images was used, however, the perceiver variability in impressions of both controlled-real and computer-generated faces was equivalent. Differences in the initial (all photos) analyses likely stem from perceivers being differentially influenced by the larger range of potentially relevant social cues available in naturalistic photographs (e.g., head tilt, angle of view, etc.; see Jenkins et al, 2011, Sutherland et al 2013 for further theoretical discussion). We find that image standardization procedures, typical in person perception research, appear to focus perceivers on a smaller set of cues when forming impressions. These procedures also appear to remove a substantial portion of perceiver variation that may be worth understanding and exploring further. Researchers should consider these advantages and limitations when selecting stimuli.

When comparing standardized real faces with computer-generated faces, target-level variation explained a greater percentage of variance in ratings of real faces. In other words, raters were more sensitive to target-variation in appearance in real than in computer-generated faces. This might be due to greater overall variability in appearance in the real face databases than in the computer-generated stimuli, or that greater detail is present in the real faces, and that this realism is influencing resulting impressions. As the realism of computer-generated faces improves with technological advances, we would expect differences between controlled-real and computer-generated faces to decrease. For now, future researchers using computer-generated faces would do well to make them as realistic as possible, and attempt to match the overall variability in their faces with the variability of real faces in their targeted population.

General Discussion

While most models of person perception have acknowledged that final impressions come from both perceiver and target characteristics, the *extent* to which perceiver and target characteristics have informed ratings of different trait impressions has remained unknown. We argue that addressing this research gap is necessary for a full understanding of the causal process of impression formation. To theoretically understand the substance and causal process of impression formation of different traits, it is important to quantify the relative contributions of perceiver- and target-characteristics, and their interaction.

Here, we have identified questions unaddressed by extant models of person perception. To what extent are different impressions driven by perceiver- and target-level factors? Do different dimensions of person perception have distinct “footprints” in perceiver- and target-level sources of variance? Do perceivers show variation in how they judge different faces? And might this vary by trait, or domain of judgment? Does the emotional extremity of a face determine the influence of the perceiver? And finally, are perceiver- and target-contributions to impressions equal for real and computer-generated faces?

To address these questions, we utilized relatively recent advances in multilevel modeling to map the extent to which perceiver and target characteristics influence final trait impressions of a large number of commonly used traits. Further, we tested specific hypotheses as to which trait impressions are more or less likely to be influenced by perceiver and target characteristics. Specifically, we calculated the perceiver- and target-ICCs as a measure of the extent to which perceiver- and target-level characteristics contribute and interact to produce trait impressions. ICCs measure the extent to which clustering of data explains variance in a dependent variable (here, impressions). Thus, an ICC approach was ideal for our purposes.

Importantly, we demonstrate that across different traits, perceiver- and target-level contributions can vary a great deal. Perceiver variability contributed from 1.8% to 36.8% of the variance in different impressions, and interaction variability 22.7% to 38.2%. Target variability contributed from 6.2% to 70.6%. Models that do not account for this variation across traits are making an *implicit* functional assumption that different traits are influenced by perceiver- and target-level characteristics to the same extent, thereby tacitly assuming that the causal process that contributes to different impressions is identical. While we believe it is unlikely that most social-cognitive researchers would make such a claim, the inflexibility of previous statistical models necessitated this assumption. The present results indicate this implicit assumption is not tenable, and is misrepresenting the rich theoretical complexity of social perception.

Our heterogeneous sample of stimuli make it likely that our conclusions are generalizable, but researchers using more specific sets of stimuli may find different patterns specific to their own sample. As discussed earlier, overall variance in the set of stimuli will influence the estimated ICCs. For instance, low variance in the attractiveness of a stimulus set yields artificially higher perceiver-ICCs (Hönekopp, 2006). Thus, it is critical to note that variance estimates are not fixed, but dependent on both characteristics of the perceivers and stimulus set.

Perceiver- and Target-Contributions to Traits and Dimensions

Which impressions are driven to a greater extent by perceiver vs. target characteristics is important for different areas of psychology examining trait impressions. A broad array of sub-disciplines examine trait judgments in their research. A key theoretical contribution of the present work is that the reported results present the first indication of what processes might drive these trait judgments. The results indicate which trait impressions arise to a greater extent from “our minds” (top of Figure 2) and which of these trait judgments are from “others’ faces” (bottom of Figure 2). We outline some specific theoretical implications of the present results as well as future directions for this work.

The current research revealed for the first time the wide variance in perceiver- and target-contributions to different traits. The results are striking in that they make clear that trait judgments that might have seemed to be somewhat similar to each other are quite different in substrate (Analysis 1). For example, though some previous work has found ratings of trustworthiness and attractiveness to be aligned (Lorenzo et al., 2010; Oosterhof & Todorov, 2008), the current results make clear they are distinct. When examining only main effects of perceiver and target variance, it appears that attractiveness is more in the eye of the beholder, whereas trustworthiness judgments are swayed by facial features (e.g., emotion; Analysis 2). Yet, when we allow for perceiver × target variance contributions (i.e., “personal taste”), we find that people show more idiosyncrasies when rating attractiveness than trustworthiness. Further, facial features contribute to a greater extent in judgments of attractiveness, thereby leading perceivers to be more consistent in rating the same face in terms of attractiveness, relative to trustworthiness (Analysis 3).

Perceiver variation affects impressions from targets to a surprising degree. In particular, there was meaningful variance driven by interplay between targets and perceivers. Perceiver × target interactions ranged from explaining 23% to 38% of the variance in ratings across dimensions: in all cases quite substantial. This varied by domain of judgment. That is, perceiver and target characteristics and their interaction contribute differentially to core dimensions of person perception (Analyses 2-3). Crucially, previous research has only examined this interaction for attractiveness (Germine et al., 2015; Hönekopp, 2006). This previous work has found that, despite a historical focus on how target characteristics influence attractiveness, around half of the variation in these impressions is actually due to idiosyncratic variation across perceivers (i.e., “personal taste”) rather than shared impressions of the target.

We note that previous work has also described personal taste as a combination of both the perceiver- and interaction-variance (Hönekopp, 2006). While both variances depend on perceiver characteristics to some extent, they exert distinct effects on ratings. The perceiver-ICC represents the extent to which one perceiver consistently rates *all* targets as higher (or lower) than another perceiver. These mean perceiver differences can be meaningful; for example, Hönekopp (2006) showed that participants who rated a set of faces as higher in attractiveness also looked at them longer, an index of reward. The interaction-ICC represents the extent to which perceivers disagree in their relative ratings of targets, and thus it depends on both targets and perceivers. For example, two friends may disagree about which film star is most attractive. Here, an attractiveness rating depends on both the perceiver and the target. Our estimates of both of these effects of personal taste agree with previous work, and help answer an age-old question by demonstrating that attractiveness is equally in the eye of the beholder (Germine et al 2015, Honekopp, 2006).

Our findings also extend previous studies by demonstrating idiosyncratic variation is relatively more important for dominance and trustworthiness dimensions than for youthful/attractiveness. While it seems intuitive to call this interaction “personal taste” for attractiveness, another way of thinking about it is: what does the trait look like to a particular perceiver? For example, what dominance “looks like” might vary across perceivers. People may have different morphological features in mind, or even be imagining different latent constructs to which different target-characteristics apply. To one person dominance may be seen as representing a large or intimidating physical appearance, to another it may be seen as displaying a confident smile.

We suggest that traits and dimensions that relate to inferences of character are more subject to idiosyncratic influences of the perceiver than traits and dimensions regarding appearance qualities (i.e. attractiveness). These results have important implications for models of social judgment that have a greater emphasis on target cues (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2016). These idiosyncrasies are not noise or error, but rather an important phenomenon in their own right, the magnitude of which will vary by domain of judgment. We encourage future researchers to allow for multiple ratings of stimuli in their designs to formally test for perceiver by target interactions (see Analysis 3).

These results also speak to the role of target-level features in core person perception judgments. Target-level characteristics contributed substantially to perceptions of faces along the dimension of trustworthiness, especially as compared to the dominance dimension. This result may be a function of the facial cues contributing to impressions of each dimension. Facial expressions of emotion are a large contributor to impressions of trustworthiness (Said et al., 2009; Zebrowitz et al., 2003). Emotional expressions may be a more salient characteristic when evaluating faces than other apparent cues to impressions of other dimensions, such as static cues like the width of the face and prominence of brow, which are important contributors to impressions of ability or dominance (Carré et al., 2010; Hehman, Flake, et al., 2015; Oosterhof & Todorov, 2008). In general, we suggest that examining how perceiver- and target-characteristics differentially contribute to dimensions of impression formation across different social categories and target attributes is critical to informing future theoretical models of social cognition.

Moderators to Perceiver- and Target-Contributions to Social Judgment

As well as examining how perceiver- and target-contributions differed by distinct traits and core social judgments, we also examined moderators of perceiver- and target-contributions to social judgment more generally. We hypothesized and found that perceivers contribute more to impressions of faces with ambiguous compared to more extreme emotional expressions (Analysis 4). As the inferences required of a perceiver increase due to ambiguity in the stimuli, so too can we expect the role of perceiver variability to drive the final impression. Importantly, these effects were found in *trait* (not emotion) judgments of targets. Typically, emotion overgeneralization is invoked to explain what leads a particular face to be more or less trusted. Yet, another logical step can be drawn. If through emotion overgeneralization, we attribute traits to faces seeming to display domain-relevant facial expressions, then extreme emotional displays may minimize perceiver contributions to traits, broadly. Thus, perhaps *posed* (as opposed to natural) emotional displays could reduce the accuracy made when making trait judgments from the face.

Finally, we also found that perceiver characteristics contribute equally to impressions of (standardized) real and computer-generated faces (Analysis 5). It is important to note that this result does not imply that there are no differences between real- and computer generated faces (e.g., Crookes et al., 2015), but only that perceiver characteristics contribute to impressions relatively equally across standardized real and computer-generated faces. Instead, a difference between these types of stimuli emerged when examining *target* characteristics, as target-level variation explained a greater percentage of variance in ratings in real than computer-generated faces. This result may be due to greater realism, detail, or variability in real faces. Future research might fruitfully use the current approach to test these possibilities.

Interestingly, perceiver variability had a larger role in impressions of naturalistic images of real faces compared to the highly controlled, standardized face photographs. This is likely because the presence of additional facial or contextual cues influenced impressions differently across different perceivers, such as the presence of jewelry or glasses, environmental information, or greater variability in facial cues such as pose, angle of photograph, and emotional expression (Hehman, Flake, et al., 2015; Sutherland et al., 2016; Todorov & Porter, 2014). There is ongoing debate as to whether and when perceivers can accurately glean person characteristics from photographs (Olivola & Todorov, 2010; Rule et al., 2013; Slepian & Ames, 2016; Todorov & Porter, 2014). Accuracy in person perception relies on “honest signals” from the target to perceivers, and the present research indicates that accuracy would be most likely observed for ratings with high target-ICCs and low perceiver-ICCs. For ratings or impressions with higher perceiver- or interaction-ICCs, variance that is *not* originating from the target would be muddying impressions. Thus, our results indicate that the context in which photographs are taken, and whether they are candid or posed, is important to consider when evaluating accuracy in person perception as they influence these ICCs.

Finally, it is worth noting that while the present research has focused on impressions of faces, the approach and results are not limited to this domain. When forming impressions, perceivers are sensitive to context, bodies, clothing, voice, and dynamic motion, among many other factors (Aviezer, Trope, & Todorov, 2012; Fessler & Holbrook, 2013b; Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011; Slepian, Young, Rutchick, & Ambady, 2013). Examining the extent to which perceiver and target characteristics contribute to impressions of these social cues is an important yet currently unexplored avenue of research, and a question that can be addressed with the present statistical approach.

Strengths

We believe the present work has several strengths. One is the scale, in that it is the largest number of ratings (n = 698,829), participants (n = 6,593), and stimuli (n = 3,353) ever used to study facial impressions, to our knowledge. This large-scale, data-driven approach was important both methodologically and theoretically. Methodologically, our estimates are more likely to be relatively stable with our large samples, and relatively unlikely to be dependent on idiosyncratic features of the photograph samples. Theoretically, others have argued that using naturally varying and heterogeneous images, such as those used here, is best to understand how impressions unfold in the real world (Burton et al., 2015; Jenkins et al., 2011; Sutherland et al., 2013). For both these reasons, the large number of stimuli from diverse sources ensures the sample is heterogeneous in its representation of different traits and representative of real world environs in which such faces are encountered, and thus has the heterogeneity necessary to allow our estimates to generalize to other samples (Hönekopp, 2006).

Similarly, in our methodological approach we implemented statistical models in which ratings were cross-classified by perceiver and target. Recent methodological work has demonstrated that aggregating ratings at either the perceiver- *or* target-level biases estimates and limits the generalizability of results (Judd et al., 2012). Accordingly, our use of cross-classified models in the current research indicates that our results should generalize beyond both our sample perceivers and targets. A final advantage of the present research is that we provide estimates of the perceiver- and target-level variance across a wide variety of commonly examined trait impressions. To our knowledge, previous research interested in quantifying perceiver and target characteristics has only recently begun, and is exclusively focused on attractiveness (Germine et al., 2015; Hönekopp, 2006). Other trait impressions, equally influential in determining important social perceptions and outcomes (Berry & Zebrowitz, 1988; Hehman, Leitner, Deegan, et al., 2013; Todorov, Mandisodza, Goren, & Hall, 2005; Wilson & Rule, 2015), have not been thoroughly and systematically quantified.

Limitations

There are several limitations of the current approach. Because the impressions involved were collected for diverse purposes and studies, they are unevenly distributed across traits. For instance, while impressions of physical strength (10.9% of sample) were regularly collected across studies, impressions of creativity (.3% of sample) were not. Estimates of perceiver- and target-ICCs will be more stable for traits with a greater number of ratings, but it is important to consider that this uneven distribution influences the stability of the estimate, and not the estimate itself, as perceiver- and target-ICCs are unrelated to the number of observations, participants, or stimuli involved in each analysis (all *p*s > .1; Analysis 1). However, the ratings of traits in the present dataset generally reflect those most commonly used in the person perception literature, and we note that, regardless of the percentage contribution of the ratings, the absolute size of the current sample (e.g. 2,020 ratings of creativity, across 101 participants and 60 stimuli; the smallest trait sample) is large enough such that all estimates are unlikely to change dramatically when examined in future work.

Further, all of the data in the present work come from one researcher (the first author). To the extent that idiosyncratic elements of the author's rating process (e.g., phrasing of instructions, computer background color, response scale wording) were consistent across the 6 years (i.e., 2011-2016) in which this data was collected, they might have systematically contributed to the results. Quantification of these ICCs by other researchers in the future will contribute to determining to what extent that might be the case. Finally, some of the analyses in the present work were exploratory and therefore we utilized a data-driven approach, as such approaches have been valuable in developing recent person perception theory (Adolphs, Nummenmaa, Todorov, & Haxby, 2016; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Therefore our results lay the initial groundwork for future research to systematically test our results in a confirmatory fashion. We have outlined many future avenues for research using our current approach.

Conclusion

In summary, the current research contributes to the person perception literature by quantifying the extent to which different trait impressions from faces are driven by perceiver and target characteristics. These results are valuable in that they can aid researchers in deciding what types of variables (perceiver or target characteristics, or interplay between the two), would predict their outcome of interest, and to what extent. In addition, these results extend theoretical models of person perception by revealing to what extent and in what contexts different impressions will be relatively driven by perceiver vs. target characteristics, revealing insight into the causal processes underlying different impressions. Estimating ICCs can offer crucial insights into specific trait impressions and the social-cognitive processes by which these impressions are formed.

By estimating and comparing ICCs, we have 1) provided greater insight into the nature of different trait impressions, 2) examined the different patterns across the different dimensions underlying person perception, 3) demonstrated a substantial effect of perceiver × target interactions in contributing to impressions, 4) revealed how emotional extremity and 5) the real vs. computer-generated source of faces is associated with the contribution of perceiver- and target-variance. Consistent across these diverse analyses, results indicate that different impressions vary a great deal in the extent to which perceiver and target characteristics contribute.

We find that trait inferences are more driven by perceiver than target characteristics, whereas impressions based on appearance qualities are more driven by target than perceiver characteristics, although all trait impressions show a greater effect of perceiver variation than hitherto considered by models of social perception. Moreover, more ambiguous stimuli are also relatively affected by perceiver variability. Finally, perceiver by target interactions are an area ripe for future research to understand how people think about and perceive these traits. Our findings demonstrate a new way to parse the variability present in trait judgments, revealing how perceivers and targets uniquely contribute to trait judgments, the interplay between the two, and how this can differ across traits.

References

Adams, R. B., Garrido, C. O., Albohn, D. N., Hess, U., & Kleck, R. E. (2016). What Facial Appearance Reveals Over Time: When Perceived Expressions in Neutral Faces Reveal Stable Emotion Dispositions. *Frontiers in Psychology*, *7*(June), 1–13. http://doi.org/10.3389/fpsyg.2016.00986

Adams, R. B., Nelson, A. J., Soto, J. a., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion*, *26*(3), 431–441. http://doi.org/10.1080/02699931.2012.666502

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150367. http://doi.org/10.1098/rstb.2015.0367

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256–274. http://doi.org/10.1037//0033-2909.111.2.256

Aviezer, H., Trope, Y., & Todorov, A. (2012). Holistic person processing: faces with bodies tell the whole story. *Journal of Personality and Social Psychology*, *103*(1), 20–37. http://doi.org/10.1037/a0027411

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*.

Beretvas, S. N. (2008). Cross-classified random effects models. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data*. Information Age publishing.

Berry, D. S., & Zebrowitz, L. A. (1988). What's in a Face? Facial Maturity and the Attribution of Legal Responsibility. *Personality and Social Psychology Bulletin*, *14*(1), 23–33. http://doi.org/0803973233

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '99*, 187–194. http://doi.org/10.1145/311535.311556

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

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2015). Identity From Variation: Representations of Faces Derived From Multiple Instances. *Cognitive Science*, *40*, 1–22. http://doi.org/10.1111/cogs.12231

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*(5), 1054–1072. http://doi.org/10.1016/j.jrp.2007.01.004

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*(10), 1194–8. http://doi.org/10.1111/j.1467-9280.2009.02423.x

Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception*, *39*(3), 356–377. http://doi.org/10.1068/p6543

Correll, J., Hudson, S. M., Guillermo, S., & Earls, H. A. (2016). Of Kith and Kin: Perceptual Enrichment, Expectancy, and Reciprocity in Face Perception. *Personality and Social Psychology Review*. http://doi.org/10.1177/1088868316657250

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. http://doi.org/10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Crookes, K., Ewing, L., Gildenhuys, J. D., Kloth, N., Hayward, W. G., Oxner, M., … Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PLoS ONE*, *10*(11), 1–18. http://doi.org/10.1371/journal.pone.0141353

Darwin, C. (1872). *The expression of the emotions in man and animals*. *The American Journal of the Medical Sciences*. London: Murray.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings. Biological Sciences / The Royal Society*, *269*(1498), 1307–12. http://doi.org/10.1098/rspb.2002.2034

DeBruine, L. M., Jones, B. C., Little, A. C., & Perrett, D. I. (2008). Social perception of facial resemblance in humans. *Archives of Sexual Behavior*, *37*(1), 64–77. http://doi.org/10.1007/s10508-007-9266-0

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. http://doi.org/10.1111/j.1467-9280.2006.01716.x

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. http://doi.org/10.1037/h0030377

Fessler, D. M. T., & Holbrook, C. (2013a). Bound to lose: physical incapacitation increases the conceptualized size of an antagonist in men. *PloS One*, *8*(8), e71306. http://doi.org/10.1371/journal.pone.0071306

Fessler, D. M. T., & Holbrook, C. (2013b). Friends shrink foes: the presence of comrades decreases the envisioned physical formidability of an opponent. *Psychological Science*, *24*(5), 797–802. http://doi.org/10.1177/0956797612461508

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. http://doi.org/10.1037//0022-3514.82.6.878

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*.

Freedman, M. B., Leary, T. F., Ossario, A. G., & Coffey, H. S. (1953). The interpersonal dimension of personality. *Journal of Personality*, *20*, 1953.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–79. http://doi.org/10.1037/a0022327

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*(1), 179–185. http://doi.org/10.1016/j.jesp.2009.10.002

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias : Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*, *27*(4), 502–517. http://doi.org/10.1177/0956797615627418

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: social status cues shape race perception. *PloS One*, *6*(9), e25107. http://doi.org/10.1371/journal.pone.0025107

Funder, D. C. (2012). Accurate Personality Judgment. *Current Directions in Psychological Science*, *21*(3), 177–182. http://doi.org/10.1177/0963721412445309

Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A. M., Smoller, J. W., Kwok, H., … Wilmer, J. B. (2015). Individual Aesthetic Preferences for Faces Are Shaped Mostly by Environments, Not Genes. *Current Biology*, *25*(20), 2684–2689. http://doi.org/10.1016/j.cub.2015.08.048

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. *Journal of Personality and Social Psychology*, *54*(6), 917–924. http://doi.org/10.1037/0022-3514.54.6.917

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827445

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early Processing of Gendered Facial Cues Predicts the Electoral Success of Female Politicians. *Social Psychological and Personality Science*, *5*(7), 815–824. http://doi.org/10.1177/1948550614534701

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and Dynamic Facial Cues Differentially Affect the Consistency of Social Evaluations. *Personality and Social Psychology Bulletin*. http://doi.org/10.1177/0146167215591495

Hehman, E., Gaertner, S. L., Dovidio, J. F., Mania, E. W., Guerra, R., Wilson, D. C., & Friel, B. M. (2012). Group status drives majority and minority integration preferences. *Psychological Science*, *23*(1), 46–52. http://doi.org/10.1177/0956797611423547

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science*, *24*(3), 289–96. http://doi.org/10.1177/0956797612451467

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2015). Picking Teams: When Dominant Facial Structure is Preferred. *Journal of Experimental Social Psychology*, *59*, 51–59. http://doi.org/10.1016/j.jesp.2015.03.007

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face-Time Continuum: Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin*. http://doi.org/10.1177/0146167214552791

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, *49*(4), 747–754. http://doi.org/10.1016/j.jesp.2013.02.015

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. http://doi.org/10.1037/0096-1523.32.2.199

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing Prejudice: Implicit Prejudice and the Perception of Facial Threat. *Psychological Science*, *14*(6), 640–643. http://doi.org/10.1046/j.0956-7976.2003.psci

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–23. http://doi.org/10.1016/j.cognition.2011.08.001

Judd, C. M., Westfall, J., & Kenny, D. a. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. http://doi.org/10.1037/a0028347

Kenny, D. a., & Albright, L. (1987). Accuracy in interpersonal perception: a social relations analysis. *Psychological Bulletin*, (3), 390–402.

Kubota, J. T., & Ito, T. a. (2007). Multiple Cues in Social Perception: The Time Course of Processing Race and Facial Expression. *Journal of Experimental Social Psychology*, *43*(5), 738–752. http://doi.org/10.1016/j.jesp.2006.10.023

Kunda, Z., Thagard, P., Kunda, Z., Thagard, P., Kintsch, W., & Ross, M. (1996). Forming Impressions From Stereotypes , Traits , and Behaviors : A Parallel-Constraint-Satisfaction Theory. *Psychological Review*, *103*(2), 284–308.

Leary, T. F. (1957). *The interpersonal diagnosis of personality; a functional theory and methodology for personality evaluation.* Eugene: Resource Publications.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 1638–1659. http://doi.org/10.1098/rstb.2010.0404

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood. Physical attractiveness and accuracy in first impressions of personality. *Psychological Science*, *21*(12), 1777–82. http://doi.org/10.1177/0956797610388048

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces-KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*. http://doi.org/10.3758/s13428-014-0532-5

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(4), 630–633. http://doi.org/10.3758/BF03206543

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, *53*(3), 431–44.

Nisbett, R. E., & Wilson, T. D. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. http://doi.org/http://dx.doi.org.ezproxy.snhu.edu/10.1037/0022-3514.35.4.250

Nunnally, J. (1978). *Psychometric Methods*. New York: McGraw-Hill.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324. http://doi.org/10.1016/j.jesp.2009.12.002

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–92. http://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133. http://doi.org/10.1037/a0014520

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Re, D. E., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2013). Facial cues to perceived height influence leadership choices in simulated war and peace contexts. *Evolutionary Psychology*, *11*(1), 89–103. http://doi.org/10.1177/147470491301100109

Re, D. E., Hunter, D. W., Coetzee, V., Tiddeman, B. P., Xiao, D., DeBruine, L. M., … Perrett, D. I. (2013). Looking like a leader-facial shape predicts perceived height and leadership ability. *PLoS ONE*, *8*(12). http://doi.org/10.1371/journal.pone.0080957

Re, D. E., Whitehead, R. D., Xiao, D., & Perrett, D. I. (2011). Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. *PLoS ONE*, *6*(3). http://doi.org/10.1371/journal.pone.0017859

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226. http://doi.org/10.1146/annurev.psych.57.102904.190208

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). FITTING THE MIND TO THE WORLD: Face Adaptation and Attractiveness Aftereffects, *14*(6), 558–566.

Rhodes, G., Morley, G., & Simmons, L. W. (2013). Women can judge sexual unfaithfulness from unfamiliar men's faces. *Biology Letters*, *9*(1), 20120908. http://doi.org/10.1098/rsbl.2012.0908

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior : Official Journal of the Human Behavior and Evolution Society*, *22*(1), 31–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11182573

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: behavioral and neural correlates. *Journal of*

Personality and Social Psychology, *104*(3), 409–26. http://doi.org/10.1037/a0031050

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264. http://doi.org/10.1037/a0014681

Secord, P. F., & Bevan, W. (1956). Personalities in Faces: III. A cross cultural comparison of impressions of physiognomy and personality in faces. *The Journal of Social Psychology*, *43*(2), 283–288. http://doi.org/10.1017/CBO9781107415324.004

Slepian, M. L., & Ames, D. R. (2016). Internalized Impressions: The Link Between Apparent Facial Trustworthiness and Deceptive Behavior Is Mediated by Targets' Expectations of How They Will Be Judged. *Psychological Science*, *27*(2), 0956797615594897-. http://doi.org/10.1177/0956797615594897

Slepian, M. L., Young, S. G., Rutchick, A. M., & Ambady, N. (2013). Quality of Professional Players' Poker Hands Is Perceived Accurately From Arm Motions. *Psychological Science*, *24*(11), 2335–2338. http://doi.org/10.1177/0956797613487384

Stephen, I. D., Law Smith, M. J., Stirrat, M., & Perrett, D. I. (2009). Facial Skin Coloration Affects Perceived Health of Human Faces. *International Journal of Primatology*, *30*(6), 845–857. http://doi.org/10.1007/s10764-009-9380-z

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, *56*(2), 249–261. http://doi.org/10.1002/dev.21183

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

Sutherland, C. A. M., Young, A. W., & Rhodes, G. (2016). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology*, 1–19. http://doi.org/10.1111/bjop.12206

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–38. http://doi.org/10.1037/a0032335

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science (New York, N.Y.)*, *308*(5728), 1623–6. http://doi.org/10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2014). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, (August 2014), 1–27. http://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Porter, J. M. (2014). Misleading First Impressions: Different for Different Facial Images of the Same Person. *Psychological Science*, (May). http://doi.org/10.1177/0956797614532474

Verkuyten, M. (2005). Ethnic Group Identification and Group Evaluation Among Minority and Majority Groups : Testing the Multiculturalism Hypothesis. *Journal of Personality and Social Psychology*, *88*(1), 121–138. http://doi.org/10.1037/0022-3514.88.1.121

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

Walker, M., & Vetter, T. (2016). Changing the Personality of a Face: Perceived Big Two and Big Five Personality Factors Modeled in Real Photographs. *Journal of Personality and Social Psychology*, *110*(4), 609–624. http://doi.org/10.1037/pspp0000064

Webster, M. A., & Macleod, D. I. A. (2011). Visual adaptation and face perception, 1702–1725. http://doi.org/10.1098/rstb.2010.0360

West, T. V, & Kenny, D. a. (2011). The truth and bias model of judgment. *Psychological Review*, *118*(2), 357–378. http://doi.org/10.1037/a0022936

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*. http://doi.org/10.1177/0956797615590992

Wolffhechel, K., Fagertun, J., Jacobsen, U. P., Majewski, W., Hemmingsen, A. S., Larsen, C. L., … Jarmer, H. (2014). Interpretation of appearance: The effect of facial features on first impressions and personality. *PLoS ONE*, *9*(9). http://doi.org/10.1371/journal.pone.0107721

Zajonc, R. b. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1. http://doi.org/10.1016/0022-1031(71)90078-3

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *7*(3), 194–215.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2007). Are Effects of Emotion Expression on Trait Impressions Mediated by Babyfaceness? Evidence From Connectionist Modeling. *Personality & Social Psychology Bulletin*, *33*(5), 648–662. http://doi.org/10.1177/0146167206297399

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*(2), 175–89. http://doi.org/10.1037/a0017990

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology*, *28*(6), 1143–1152. http://doi.org/10.1037//0012-1649.28.6.1143

Zebrowitz, L. A., & Montepare, J. M. (2005). Psychology. Appearance DOES matter. *Science (New York, N.Y.)*, *308*(5728), 1565–6. http://doi.org/10.1126/science.1114170

Footnotes

[1]Most studies for which these data were collected were not interested in racial demographics, and this information was not consistently obtained.

Supplementary Materials

R code to run and derive ICCs from cross-categorized multilevel models

The below can be fully copy and pasted into R. Comments are denoted with a '#', and anything without a # should be run as code.

```
#Downloading and installing lme4, an R package for multilevel modeling

install.packages("lme4")
library(lme4)

#Building the cross-classified model
#First we present a basic model
#In which: m = the created model, dv = dependent variable, c1 = cluster 1, c2 = cluster 2,
# and dataset = dataset name
#Thus, dv is cross-categorized by both c1 and c2.

m <- lmer(dv ~ 1 + (1 | c1) + (1 | c2), data=dataset)

#Here is an example model in the context of Hehman Sutherland Flake and Slepian to illustrate
#Friendly is cross-categorized by both rater and target

m <- lmer(friendly ~ 1 + (1 | rater) + (1 | target), data=dataset)

#Analyses for calculating ICCs
#The below command will return, in part, a section labeled 'random effects'. We provide an example

summary(m)

#Random effects:
#Groups      Name          Variance     Std.Dev.
#rater       (Intercept)   0.5868       0.7660
#target      (Intercept)   0.2038       0.4514
#Residual                  1.7427       1.3201

#The first two bolded numbers above represent the taus (τ) for rater and target, respectively. The bottom
#bolded value (labeled residual) is sigma squared (σ²) in the ICC equation in the paper.

#To calculate rater ICC
#ICC = rater tau / (sigma squared + rater tau + target tau)
#To calculate target ICC
#ICC = target tau / (sigma squared + rater tau + target tau)
#The resulting numbers represent the percentage of variance from between-rater (rater-ICC) or between-
#target (target-ICC), respectively.

#To run models that can estimate interactions between clusters use the below code
#A minimum of two observations per c1 and c2 is required for this model to run

m <- lmer(dv ~ 1 + (1 | c1) + (1 | c2) + (1 | c1:c2), data=dataset)
```

Supplementary Figure 1

Supplementary Figure 1. Examples of five levels of emotional display in stimuli using in Analysis 4, ranging from subtly angry to subtly happy. These five levels were recoded into three levels of neutral to most extreme.

Correlation Matrix

Correlations between trait ratings (averaged across all participants for each stimulus) used to form dimensions in Analysis 2.
Thus, *N* reflects number the number of stimuli for which the two traits were each collected. The large degree of variability in *N* is due to different trait ratings and different stimuli being included across different studies.

		Aggressive	Caring	Friendly	Likeable	Trustworthy	Warm	Wise	Assertive	Competent	Dominant
Aggressive	Pearson Correlation	1	-.711**	-.714**	-.486**	-.630**	-.488**	-.270**	.627**	-.312**	.644**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.003	.000
	N	314	87	146	87	123	219	191	87	87	87
Caring	Pearson Correlation	-.711**	1	.891**	.811**	.819**	.863**	.b	-.188	.717**	-.241*
	Sig. (2-tailed)	.000		.000	.000	.000	.000		.080	.000	.024
	N	87	87	87	87	87	87	0	87	87	87
Friendly	Pearson Correlation	-.714**	.891**	1	.789**	.818**	.935**	.478**	-.197	.619**	-.551**
	Sig. (2-tailed)	.000	.000		.000	.000	0.000	.000	.067	.000	.000
	N	146	87	1729	87	977	859	59	87	650	1355
Likeable	Pearson Correlation	-.486**	.811**	.789**	1	.814**	.815**	.b	.039	.884**	-.108
	Sig. (2-tailed)	.000	.000	.000		.000	.000		.723	.000	.320
	N	87	87	87	87	87	87	0	87	87	87
Trustworthy	Pearson Correlation	-.630**	.819**	.818**	.814**	1	.838**	.b	-.597**	.619**	-.655**
	Sig. (2-tailed)	.000	.000	.000	.000		.000		.000	.000	.000
	N	123	87	977	87	1902	859	0	633	480	1866
Warm	Pearson Correlation	-.488**	.863**	.935**	.815**	.838**	1	.103	-.034	.519**	-.731**
	Sig. (2-tailed)	.000	.000	0.000	.000	.000		.242	.756	.000	.000
	N	219	87	859	87	859	1243	132	87	339	859
Wise	Pearson Correlation	-.270**	.b	.478**	.b	.b	.103	1	.b	.b	.b
	Sig. (2-tailed)	.000		.000			.242				
	N	191	0	59	0	0	132	191	0	0	0
Assertive	Pearson Correlation	.627**	-.188	-.197	.039	-.597**	-.034	.b	1	-.024	.819**
	Sig. (2-tailed)	.000	.080	.067	.723	.000	.756			.777	.000
	N	87	87	87	87	633	87	0	633	141	633
Competent	Pearson Correlation	-.312**	.717**	.619**	.884**	.619**	.519**	.b	-.024	1	-.131**
	Sig. (2-tailed)	.003	.000	.000	.000	.000	.000		.777		.001
	N	87	87	650	87	480	339	0	141	1371	692
Dominant	Pearson Correlation	.644**	-.241*	-.551**	-.108	-.655**	-.731**	.b	.819**	-.131**	1
	Sig. (2-tailed)	.000	.024	.000	.320	.000	.000		.000	.001	
	N	87	87	1355	87	1866	859	0	633	692	2244
Mean	Pearson Correlation	.b	.b	-.777**	.b	.b	.b	.b	.b	.b	.b
	Sig. (2-tailed)			.000							
	N	0	0	60	0	0	0	0	0	0	0
Physically Powerful	Pearson Correlation	.478**	.b	.144	.b	.b	.b	-.043	.b	.b	.b
	Sig. (2-tailed)	.000		.276				.744			
	N	59	0	59	0	0	0	59	0	0	0
Strong	Pearson Correlation	.326**	.001	-.212**	.074	-.411**	-.378**	.b	.649**	.281**	.678**
	Sig. (2-tailed)	.002	.995	.000	.499	.000	.000		.000	.000	.000
	N	87	87	1463	87	881	862	0	87	570	1255
Socially Powerful	Pearson Correlation	.401**	.b	.020	.b	.b	.b	.222	.b	.b	.b
	Sig. (2-tailed)	.002		.881				.091			
	N	59	0	59	0	0	0	59	0	0	0
Attractive	Pearson Correlation	-.252*	.572**	.313**	.665**	.412**	.468**	.b	-.098*	.468**	-.059*
	Sig. (2-tailed)	.019	.000	.000	.000	.000	.000		.013	.000	.025
	N	87	87	872	87	1072	339	0	633	1310	1446
Creative	Pearson Correlation	.b	.b	.535**	.b	.b	.b	.b	.b	.b	.b
	Sig. (2-tailed)			.000							
	N	0	0	60	0	0	0	0	0	0	0
Healthy	Pearson Correlation	-.197	.520**	.493**	.530**	.477**	.538**	.b	.292**	.650**	.281**
	Sig. (2-tailed)	.067	.000	.000	.000	.000	.000		.006	.000	.008
	N	87	87	87	87	87	87	0	87	87	87
Intelligent	Pearson Correlation	-.345**	.648**	.510**	.752**	.650**	.492**	.b	.179	.771**	.065
	Sig. (2-tailed)	.001	.000	.000	.000	.000	.000		.098	.000	.233
	N	87	87	604	87	87	339	0	87	694	336
Smart	Pearson Correlation	-.219*	.618**	.579**	.813**	.682**	.634**	.b	.327**	.912**	.165
	Sig. (2-tailed)	.041	.000	.000	.000	.000	.000		.002	.000	.126
	N	87	87	87	87	87	87	0	87	87	87
Youthful	Pearson Correlation	-.247*	.103	.100	.258*	.090	.031	.b	-.371**	.004	-.496**
	Sig. (2-tailed)	.021	.340	.357	.016	.406	.775		.000	.971	.000
	N	87	87	87	87	87	87	0	87	87	87

		Mean	Physically Powerful	Strong	Socially Powerful	Attractive	Creative	Healthy	Intelligent	Smart	Youthful
Aggressive	Pearson Correlation	.[b]	.478**	.326**	.401**	-.252*	.[b]	-.197	-.345**	-.219*	-.247*
	Sig. (2-tailed)		.000	.002	.002	.019		.067	.001	.041	.021
	N	0	59	87	59	87	0	87	87	87	87
Caring	Pearson Correlation	.[b]	.[b]	.001	.[b]	.572**	.[b]	.520**	.648**	.618**	.103
	Sig. (2-tailed)			.995		.000		.000	.000	.000	.340
	N	0	0	87	0	87	0	87	87	87	87
Friendly	Pearson Correlation	-.777**	.144	-.212**	.020	.313**	.535**	.493**	.510**	.579**	.100
	Sig. (2-tailed)	.000	.276	.000	.881	.000	.000	.000	.000	.000	.357
	N	60	59	1463	59	872	60	87	604	87	87
Likeable	Pearson Correlation	.[b]	.[b]	.074	.[b]	.665**	.[b]	.530**	.752**	.813**	.258*
	Sig. (2-tailed)			.499		.000		.000	.000	.000	.016
	N	0	0	87	0	87	0	87	87	87	87
Trustworthy	Pearson Correlation	.[b]	.[b]	-.411**	.[b]	.412**	.[b]	.477**	.650**	.682**	.090
	Sig. (2-tailed)			.000		.000		.000	.000	.000	.406
	N	0	0	881	0	1072	0	87	87	87	87
Warm	Pearson Correlation	.[b]	.[b]	-.378**	.[b]	.468**	.[b]	.538**	.492**	.634**	.031
	Sig. (2-tailed)			.000		.000		.000	.000	.000	.775
	N	0	0	862	0	339	0	87	339	87	87
Wise	Pearson Correlation	.[b]	-.043	.[b]	.222	.[b]	.[b]	.[b]	.[b]	.[b]	.[b]
	Sig. (2-tailed)		.744		.091						
	N	0	59	0	59	0	0	0	0	0	0
Assertive	Pearson Correlation	.[b]	.[b]	.649**	.[b]	-.098*	.[b]	.292**	.179	.327**	-.371**
	Sig. (2-tailed)			.000		.013		.006	.098	.002	.000
	N	0	0	87	0	633	0	87	87	87	87
Competent	Pearson Correlation	.[b]	.[b]	.281**	.[b]	.468**	.[b]	.650**	.771**	.912**	.004
	Sig. (2-tailed)			.000		.000		.000	.000	.000	.971
	N	0	0	570	0	1310	0	87	694	87	87
Dominant	Pearson Correlation	.[b]	.[b]	.678**	.[b]	-.059*	.[b]	.281**	.065	.165	-.496**
	Sig. (2-tailed)			.000		.025		.008	.233	.126	.000
	N	0	0	1255	0	1446	0	87	336	87	87
Mean	Pearson Correlation	1	.[b]	.[b]	.[b]	-.252	-.302*	.[b]	.935**	.[b]	.[b]
	Sig. (2-tailed)					.052	.019		.000		
	N	60	0	0	0	60	60	0	60	0	0
Physically Powerful	Pearson Correlation	.[b]	1	.[b]	.634**	.[b]	.[b]	.[b]	.[b]	.[b]	.[b]
	Sig. (2-tailed)				.000						
	N	0	59	0	59	0	0	0	0	0	0
Strong	Pearson Correlation	.[b]	.[b]	1	.[b]	.408**	.[b]	.613**	.083*	.243*	-.261**
	Sig. (2-tailed)					.000		.000	.036	.023	.005
	N	0	0	1553	0	699	0	87	634	87	114
Socially Powerful	Pearson Correlation	.[b]	.634**	.[b]	1	.[b]	.[b]	.[b]	.[b]	.[b]	.[b]
	Sig. (2-tailed)		.000								
	N	0	59	0	59	0	0	0	0	0	0
Attractive	Pearson Correlation	-.252	.[b]	.408**	.[b]	1	.318*	.728**	.438**	.604**	.311**
	Sig. (2-tailed)	.052		.000			.013	.000	.000	.000	.001
	N	60	0	699	0	2267	60	87	883	87	114
Creative	Pearson Correlation	-.302*	.[b]	.[b]	.[b]	.318*	1	.[b]	-.252	.[b]	.[b]
	Sig. (2-tailed)	.019				.013			.052		
	N	60	0	0	0	60	60	0	60	0	0
Healthy	Pearson Correlation	.[b]	.[b]	.613**	.[b]	.728**	.[b]	1	.644**	.557**	.162
	Sig. (2-tailed)			.000		.000			.000	.000	.134
	N	0	0	87	0	87	0	87	87	87	87
Intelligent	Pearson Correlation	.935**	.[b]	.083*	.[b]	.438**	-.252	.644**	1	.808**	.058
	Sig. (2-tailed)	.000		.036		.000	.052	.000		.000	.542
	N	60	0	634	0	883	60	87	943	87	114
Smart	Pearson Correlation	.[b]	.[b]	.243*	.[b]	.604**	.[b]	.557**	.808**	1	-.024
	Sig. (2-tailed)			.023		.000		.000	.000		.827
	N	0	0	87	0	87	0	87	87	87	87
Youthful	Pearson Correlation	.[b]	.[b]	-.261**	.[b]	.311**	.[b]	.162	.058	-.024	1
	Sig. (2-tailed)			.005		.001		.134	.542	.827	
	N	0	0	114	0	114	0	87	114	87	114

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).
b Cannot be computed because at least one of the variables is constant.