PERSON PERCEPTION ACROSS GROUP BOUNDARIES:

A DYNAMIC MODEL OF PERCEPTION ACROSS RACE AND GENDER LINES

by

Sally Yuan-Jun Xie,

Honours Bachelor of Science, University of Toronto, 2016

A thesis

presented to Ryerson University

in partial fulfillment of the requirements for the

degree of

Master of Arts

in the Program of

Psychology

Toronto, Ontario, Canada, 2018

© (Sally Yuan-Jun Xie) 2018

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Person Perception across Group Boundaries: A Dynamic Model of Perception across Race and Gender Lines

Sally Yuan-Jun Xie

Master of Arts, 2018

Psychology, Ryerson University

Abstract

People form impressions of others from their faces, inferring character traits (e.g., friendly) along two broad, influential dimensions: Warmth and Competence. Although these two dimensions are presumed to be independent, research has yet to examine the generalizability of this model to cross-group impressions, despite extant evidence that Warmth and Competence are *not* independent for outgroup targets. This thesis explores this possibility by testing models of person perception for *own-group* and *other-group* perceptions, implementing confirmatory factor analysis in a structural equation modeling framework, and analyzing the underlying trait space using representational similarity analysis. I fit 402,473 ratings of 873 unique faces from 5,040 participants on 14 trait impressions to own-group and other-group models, exploring whether perceptions across race and gender are more unidimensional. Results indicate that current models of face perception fit poorly and are not universal as presumed: the space of trait impressions varies depending on targets' race and gender.

Keywords: person perception, impression formation, face perception, intergroup processes, social cognition

Acknowledgements

I would like to thank my brilliant and generous advisor, Dr. Eric Hehman, for his guidance on a project of great interest to me. Working with Eric is inspiring, and I'm constantly reminded of how much more I have to learn about life, research, and powerlifting. As a mentor, he is lovely and thoughtful, and a delight to be around – I'm grateful for the time and unwavering support he has provided throughout my graduate training.

Tremendous gratitude goes toward my examining committee, Dr. Becky Choma, Dr. Margaret Moulson, and Dr. Stephen Want, for their insight, expertise, and kind words of encouragement. I'm delighted to see this thesis improve through their feedback. A special thank-you also goes out to Dr. Jessica Flake, who spent many hours ensuring I understood enough about structural equation models to write about them semi-intelligibly.

Thank you to my cherished friends, colleagues, and lab mates at Ryerson for two memorable years. I will miss being around such a wonderful community and the warm, collaborative culture of mutual support and free food that they have cultivated.

Finally, my deepest gratitude goes out to my family and friends for always being there for me and feigning interest in my research. I dedicate this 100-page thesis to them because I strongly believe they will read the whole thing.

Table of Contents

Page

Author's Declaration ii
Abstract iii
Acknowledgements iv
List of Tables vii
List of Figures viii
List of Appendices ix
Introduction 1
Social Impressions from Faces 4
Social Impressions from Faces are Driven by Morphological Features 7
Data-driven methods 11
Major Dimensions Underlying Person Perception 12
Warmth/Trustworthiness and Competence/Dominance 14
Dimensions of social cognition 14
Dimensions extracted from faces 15
Youthful-attractiveness dimension 18
Generalizability of the Two-dimensional Model 19
Intergroup Effects in Face Perception 21
Activation of stereotype-congruent impressions 23
Contextual influences on perceptions of threat 25
The Present Study 28
Method 30
Analytic Approach 30
Representational similarity analysis 32
Source of the Data 34
Participants 36
Stimuli 37

Results 38
Two-Dimensional Models 38
Model fit 40
Three-Dimensional Models 44
Model fit 46
Representational Similarity Analysis 48
Descriptive statistics 51
Within-subjects effects 51
General Discussion 52
Classic Models of Person Perception 56
Competence vs. dominance 59
Representational Similarity of the Social Trait Space Differs Across Groups 60
Limitations and Future Directions 64
Conclusion 67
References 82

List of Tables

Table		Page
1	Indices of model fit for the two-dimensional model of face perception across own-group and cross-group perceptions	41
2	Parameter estimates for the two-dimensional model of White participants' perceptions of female and male White, Black, and Asian faces	43
3	Indices of model fit for the three-dimensional model of face perception across own-group and cross-group perceptions	46
4	Parameter estimates for the three-dimensional model of White participants' perceptions of female and male White, Black, and Asian faces	47
5	Estimated Marginal Means from an RMANOVA on the 6 trait-pair correlation matrices in a 2 (Target Gender) × 3 (Target Race) design	51

Supplementary Table		Page
1	Indices of model fit for an exploratory one-factor model of face perception across own-group and cross-group perceptions	77
2	Parameter estimates for the unidimensional model of White participants' perceptions of female and male White, Black, and Asian faces	78

List of Figures

Figure Page

1 Independent and correlated models of person perception, and their corresponding statistical models. The dimensions of warmth and competence are independent in the (a) classic model and correlated in the (b) proposed model 2

2 Example stimuli 36

3 SEM path diagram for a confirmatory factor analysis, representing the loadings of 14 traits onto their latent dimensions of Trustworthiness and Dominance, which were fixed to be orthogonal. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait (e.g., trustworthy) served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable 39

4 SEM path diagram for a three-dimensional CFA, representing the loadings of 14 traits onto their latent dimensions of Trustworthiness, Dominance, and Youthful/Attractiveness, which were fixed to be orthogonal. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait (e.g., trustworthy) served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable 45

5 Correlation matrices of 14 traits for White participants' perceptions of female and male White, Black, and Asian faces. Darker colours represent stronger correlations (closer to r = 1 or r = -1) between pairs of traits. Blank squares indicate trait-pairs that are non-significant at $p < .05$ 50

Supplementary Figures Page

1 SEM path diagram for a confirmatory factor analysis, representing the loadings of 14 traits onto a single general factor. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable 76

2 Density plots showing the distributions of 14 trait ratings for each social group. Each distribution represents White participants' ratings of White, Black, or Asian female or male targets (e.g., W.WF = White perceivers' ratings of White female targets).....79

List of Appendices

Appendix Page

Appendix A Study Materials 69

Appendix A.1 Consent Form on Amazon Mechanical Turk 69

Appendix A.2 Sample Study Portal on Website 71

Appendix A.3 Sample Rating Task 72

Appendix A.4 Demographic Information 73

Appendix A.5 Manipulation Check 74

Appendix A.6 Example Stimuli 75

Appendix B Supplementary Analyses 76

Appendix B.1 One-Factor Model in CFA 76

Appendix B.2 Density Plots of Ratings on Each Trait 79

Introduction

People regularly form impressions of others based on their facial appearance, drawing social inferences about their character (e.g., trustworthy) through a perceptual process that is effortless and spontaneous (Oosterhof & Todorov, 2008; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Willis & Todorov, 2006). There are as many ways to evaluate people as there are adjectives in the dictionary, but most of these trait evaluations are closely related – individuals perceived as friendly are likely to also be perceived as trustworthy. Past researchers have used data-reduction techniques to distill these trait impressions down to two latent dimensions underlying person perception, which reflect exigent social pressures to quickly distinguish friend from foe: (1) whether a person has good or ill intentions: *Warmth*, and (2) that person's ability to enact those intentions: *Competence* (Fiske, Cuddy, & Glick, 2007; Freeman & Ambady, 2011; Todorov, Said, Engell, & Oosterhof, 2008). These two-dimensional models of group and person perception have been extremely influential, with research showing that social evaluations along these dimensions effect a panoply of downstream societal consequences, from election results (Ballew & Todorov, 2007; Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014; Olivola & Todorov, 2010a; Todorov, Mandisodza, Goren, & Hall, 2005) to judicial outcomes (Blair, Judd, & Chapleau, 2004; Wilson & Rule, 2015) to dating and hiring decisions (Stevenage & McKay, 1999).

These models, with their two independent dimensions of Warmth and Competence, are theoretically presumed to be universal, reflecting how humans perceive and evaluate all other humans. Critically, however, these models were derived almost exclusively from White participants' evaluations of White male targets (Hehman, Flake, & Freeman, 2015; Oosterhof & Todorov, 2008; Todorov et al., 2015). The modern world is comprised of numerous groups (e.g.,

across race, gender, sexuality), and some of the most consequential impressions are those formed across group boundaries. Yet the generalizability of this prominent model of person perception has not been scrutinized, despite extant research indicating that the dimensions of Warmth and Competence are *not* independent for cross-group impressions. Competent women are perceived as less warm (Cuddy et al., 2009; Eagly & Kite, 1987; Glick & Fiske, 1996); physically-dominant Black men as more ill-intentioned (Livingston & Pearce, 2009; Wilson, Hugenberg, & Rule, 2017). For these targets, perceptions along each dimension may be correlated or occur in a more "unidimensional" manner (Figure 1) in which judgements on one dimension bias the other.

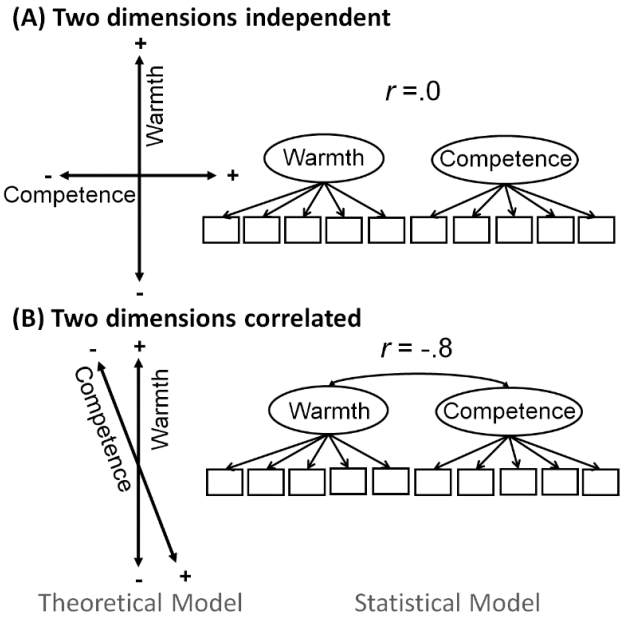


Figure 1. Independent and correlated models of person perception, and their corresponding statistical models. The dimensions of warmth and competence are independent in the (a) classic model and correlated in the (b) proposed model.

There are several reasons to believe that impressions along the dimensions of Warmth and Competence may be correlated for cross-group impressions. In a complex social world that requires us to constantly and rapidly evaluate those around us (friend or foe, approach or avoid,

help or harm), efficient impression formation confers many advantages (Ambady & Rosenthal, 1992; Zebrowitz & Montepare, 2008). To streamline the process, we categorize others into groups (Macrae & Bodenhausen, 2000), infer their intentions and their ability to enact those intentions (Mason, Tatkow, & Macrae, 2005), and adjust our behaviour accordingly. Social-cognitive models, which emphasize the category salience of outgroup members, find that merely categorizing people into groups alters the type or depth of processing that outgroup targets receive, such that cross-group perceptions are influenced by a host of top-down cognitive and motivational factors (e.g., stereotype content, evaluative context, threat management; Fiske et al., 2007; Fiske, Cuddy, Glick, & Xu, 2002; Hehman, Mania, & Gaertner, 2010; Hugenberg & Bodenhausen, 2003; Zebrowitz & Montepare, 2008). We may not always have the luxury of perceiving outgroup members as multidimensional, complex individuals. Evolutionary theories of adaptive threat management suggest that to minimize risk (Bar, Neta, & Linz, 2006; Fiske et al., 2007; Haselton & Nettle, 2006; Holbrook et al., 2014), it may be more adaptive to perceive both negatively-intentioned and high-ability targets as threatening, without evaluating them on additional dimensions (e.g., if a person is perceived as high in ability, without additional information, it is safer to infer negative intentions; Holbrook, Fessler, & Navarrete, 2016). Thus, it may have been functionally adaptive to associate high-Competence outgroup members with low Warmth (and vice versa) and avoid them accordingly, engendering a predisposition to perceive outgroup targets in a correlated, unidimensional manner (Fessler & Holbrook, 2013a; Fiske et al., 2007, 2002; Holbrook et al., 2016; Oosterhof & Todorov, 2008).

This thesis explores the possibility that dimensions of person perception are fluid across group boundaries. By extending the classic two-dimensional model, I test the hypotheses that dimensions of person perception are more or less correlated for perceptions across social

categories, both as a function of gender and race. The goal of the current research is to generalize models of person perception to all individuals, informing our understanding of how such impressions are formed. Accordingly, I begin with a brief review of the facial first impressions literature, highlighting the considerable influence of social impressions on real-world outcomes, as well as the features that give rise to these impressions, before turning to the development of current dominant models of person perception.

Social Impressions from Faces

Faces are a powerfully salient source of information for social attributions, informing our theories and expectations about other people's mental states and behaviours (Zebrowitz, 1997; Zebrowitz & Montepare, 2008). Within 100 milliseconds of exposure to a face, perceivers form specific trait impressions (e.g., trustworthy, aggressive) of the target that are stable and do not change with longer exposure time (Ballew & Todorov, 2007; Bar et al., 2006; Borkenau, Brecke, Möttig, & Paelecke, 2009; Mason et al., 2005; Rule & Ambady, 2008a; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Even if such inferences may not be accurate (for a review, see Todorov et al., 2015), there is consensus in these trait impressions across different observers (Todorov, Baron, & Oosterhof, 2008; Willis & Todorov, 2006) and cultures (Albright et al., 1997; Berry, Wero, & Julia, 1993; Rule et al., 2010), and even among young children (Cogsdill, Todorov, Spelke, & Banaji, 2014) – suggesting that the propensity to infer traits from faces is not a cultural by-product but arises out of innate processes. For instance, perceptions of babyfaced adults show high inter-rater agreement on a constellation of traits (e.g., physical strength, dominance, intelligence, trustworthiness) across cultures and age groups (Berry & Zebrowitz, 1988; McArthur & Apatow, 1984; McArthur & Berry, 1987; Montepare &

Zebrowitz-McArthur, 1989; Zebrowitz & Montepare, 1992), indicating that impressions from faces of these traits are robust and reliable.

Regardless of accuracy, the impressions gleaned from faces are pervasive and consequential, predicting real-world outcomes across a wide variety of domains – often in ways that are systematically biased. Within politics, voting decisions and electoral success are predicted by candidates' facial competence: the more competent-looking their faces are, the more votes they receive and the more likely they are to win (Ballew & Todorov, 2007; Hehman, Carpinella, et al., 2014; Lenz & Lawson, 2011; Little, Burriss, Jones, & Roberts, 2007; Olivola & Todorov, 2010a; Todorov et al., 2005), and this effect has been replicated for cross-cultural judgements of competence (Lawson, Lenz, Baker, & Myers, 2010). Other trait inferences predictive of electoral success include dominance, which is negatively linked to success (although this relationship was moderated by peacetime versus wartime contexts; Chen, Jing, & Lee, 2014; Little et al., 2007; Re, DeBruine, Jones, & Perrett, 2013), gender typicality (faces perceived as gender-atypical and less feminine received fewer votes, but only for female politicians in conservative areas; Carpinella, Hehman, Freeman, & Johnson, 2015; Carpinella & Johnson, 2013; Hehman, Carpinella, et al., 2014), sociability (Castelli, Carraro, Ghitti, & Pastore, 2009), and threat (Mattes et al., 2010).

Beyond politics, social inferences from faces also inform who we accept, date, and reward. When forming groups, facial dominance predicts selection for group membership in physically-competitive contexts and when intergroup conflict is salient (Hehman, Leitner, Deegan, & Gaertner, 2015). When selecting a mate, perceived facial extraversion positively predicts online dating success for men, whereas facial competence negatively predicts success for women, even after controlling for attractiveness (Olivola et al., 2014, as cited in Todorov et

al., 2015). Within the business domain, CEOs with more competent and dominant faces are hired by higher-status companies and receive larger salaries (Graham, Harvey, & Puri, 2016; Rule & Ambady, 2008b, 2009), irrespective of actual performance (Graham et al., 2016). Other studies have found this effect to be moderated by gender and race, with only a marginal effect of facial competence for female CEOs and an inverse relationship for Black CEOs (Livingston & Pearce, 2009; Pillemer, Graham, & Burke, 2014) – instead, facial warmth, not competence, predicted better outcomes for both groups.

Of particular concern are the social inferences that have consequences within the legal system. Experimental studies have demonstrated that impressions from facial appearance can predict real-world judgements of guilt and sentencing severity. Defendants with untrustworthy-looking (Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015) or stereotype–crime congruent faces are more likely to be found guilty, even when the evidence against them is less compelling (Dumas & Testé, 2006; Macrae & Shepherd, 1989; Porter et al., 2010; Shoemaker, South, & Lowe, 1972). In addition, having Afrocentric (i.e., stereotypically Black) facial features predicts harsher sentences (Blair et al., 2004) and a greater likelihood of receiving the death sentence (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006), even after controlling for inmates' actual race – reflecting feature-based racial discrimination. Finally, babyfaced defendants are more likely to be exonerated in civil cases when charged with crimes requiring intent versus crimes of negligence (Berry & Zebrowitz, 1988; Zebrowitz & McDonald, 1991), and babyfaced plaintiffs receive greater monetary rewards when defendants look facially mature (Zebrowitz & McDonald, 1991).

These findings provide compelling evidence that social impressions gleaned from faces have downstream consequences for society, influencing our decision-making even in situations

that ought to be guided by more objective information. Furthermore, the impact of a particular facial trait impression varies with context (e.g., business versus juridical settings), target characteristics (e.g., race and ethnicity), and perceiver characteristics (e.g., politically informed). Critically, although the influence of face-based inferences is sometimes attenuated by access to more relevant information (such as political knowledge), the biases that arise from these initial social impressions are never entirely eliminated (Olivola & Todorov, 2010b).

Social Impressions from Faces are Driven by Morphological Features

Given that social inferences from faces are so pervasive and consequential, and not necessarily accurate, how and why do we constantly infer trait impressions from faces? Modern approaches to this question have focused on identifying the facial cues (e.g., mouth curvature, skin colouration, etc.) that are important for social impressions, and examining these cues in theoretical frameworks that emphasize the functional significance of faces as sources of social information (Zebrowitz-McArthur & Baron, 1983; Zebrowitz, 1997; for a review, see Zebrowitz & Montepare, 2008). According to the ecological approach (Berry et al., 1993; Zebrowitz-McArthur & Baron, 1983; Zebrowitz & Montepare, 2008), faces convey adaptive information about the social interactions that they afford, providing visual indicators of a person's attention (Haxby, Hoffman, & Gobbini, 2000; Mason et al., 2005), intent (Oosterhof & Todorov, 2008), social identity (Freeman & Ambady, 2011; Freeman, Pauker, Apfelbaum, & Ambady, 2010; Hehman, Ingbretsen, & Freeman, 2014; Kubota & Ito, 2007), and emotional state (Adams, Nelson, Soto, Hess, & Kleck, 2012; Bruce & Young, 1986; Darwin, 1872; Ekman & Friesen, 1971). For example, the face of a baby – cute, warm, weak, incompetent – elicits approach and affiliative responses (McArthur & Berry, 1987; Zebrowitz, 1997; Zebrowitz & Montepare,

1992), whereas an angry face – signalling threat and ill intentions – elicits avoidance and defensive responses (Marsh, Adams, & Kleck, 2005; Marsh, Ambady, & Kleck, 2005).

Although our perceptions of faces will sometimes be accurate, facilitating an appropriate response to the immediate situation, it may be inappropriate to generalize transient emotional cues (a smile) to stable trait inferences congruent with these cues (a friendly person). Yet people readily extrapolate enduring personality traits from momentary emotional states (Secord, Dukes, & Bevan, 1954). This temporal extension bias (Secord, 1958) is at the core of *overgeneralization effects*, which posit that social inferences from faces are driven by a tendency to overgeneralize adaptive facial cues (e.g., emotions, social identity) to stable impressions about people whose facial appearance resembles these cues, producing biased perceptions (Oosterhof & Todorov, 2008; Todorov et al., 2015; Zebrowitz, 1997; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003; Zebrowitz & Montepare, 2008). Specifically, the emotion overgeneralization theory proposes that certain facial features are so useful in guiding interpersonal behaviour that we are naturally predisposed to detecting them. Consequently, even minor resemblances to emotional expressions in the form of static facial structure and variations in musculature are misattributed to stable trait impressions (Adams et al., 2012; Oosterhof & Todorov, 2009; Said, Sebe, & Todorov, 2009; Secord, 1958; Secord et al., 1954; Zebrowitz, Kikuchi, & Fellous, 2010; Zebrowitz & Montepare, 2008). For example, a person with downturned eyebrows, a classic sign of an angry expression, may be evaluated as more dominant (Adams et al., 2012; Keating, Mazur, & Segall, 1981), whereas larger eyes and rounder faces resembling those of a baby are perceived as warm and honest (Zebrowitz, 1997; Zebrowitz & Montepare, 1992).

The overgeneralization hypothesis has received widespread empirical support in the face perception literature. Research by Zebrowitz and colleagues find that a babyfaced appearance

influences social impressions, and that babyfaced adults are perceived as warm, honest, weak, submissive, and naïve (McArthur & Apatow, 1984; McArthur & Berry, 1987; Montepare & Zebrowitz-McArthur, 1989; Zebrowitz et al., 2003; Zebrowitz, Kikuchi, & Fellous, 2007; Zebrowitz & Montepare, 1992). These traits are predicted not only by the presence of babyish features such as a round face and large eyes, but also by connectionist models examining the structural similarity between adult and baby faces (Zebrowitz et al., 2003).

Similar findings have been obtained with other impressions such as attractiveness, which correlates positively with trait inferences on a host of other dimensions, including competence, intelligence, sociability, and health (the attractiveness halo effect; Eagly, Ashmore, Makhijani, & Longo, 1991; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Zebrowitz, Hall, Murphy, & Rhodes, 2002; Zebrowitz & Rhodes, 2004). The facial characteristics that predict perceptions of attractiveness may reflect evolutionary indicators of fitness, and include symmetry (which signals developmental stability; Scheib, Gangestad, & Thornhill, 1999), averageness (genetic diversity; Thornhill & Gangestad, 1999), skin colouration (health; Re, Whitehead, Xiao, & Perrett, 2011; Stephen, Law Smith, Stirrat, & Perrett, 2009), and sexual dimorphism (sexual maturity; Folstad & Karter, 1992). Furthermore, these facial cues independently predict other trait impressions that correlate with attractiveness – facial asymmetry and distinctiveness are perceived as less intelligent and healthy (Rhodes et al., 2001; Zebrowitz et al., 2002; Zebrowitz & Rhodes, 2004), and women with high cheekbones are perceived as more intelligent, healthy, and sociable (Cunningham, Roberts, Barbee, Druen, & Wu, 1995).

Critically, there is extensive evidence that social inferences also arise from the overgeneralization of emotional expressions. Consistent with the temporal extension bias, facial expressions which signal transient affective states are overgeneralized to impressions of stable

personality traits (Montepare & Dobish, 2003; Secord, 1958; Zebrowitz & Montepare, 2008). For instance, happy, smiling faces are perceived as more trustworthy (Oosterhof & Todorov, 2009), whereas angry faces are perceived as more dominant (Hess, Blairy, & Kleck, 2000; Knutson, 1996; Montepare & Dobish, 2003). More importantly, this pattern extends to emotionally neutral faces. Facial features that resemble specific emotional expressions, owing to variations in facial musculature (Hehman, Flake, et al., 2015; Sutherland, Young, & Rhodes, 2016; Todorov & Porter, 2014) and static morphological features (Hehman, Leitner, & Gaertner, 2013; Marsh, Adams, et al., 2005; Oosterhof & Todorov, 2009; Posamentier & Abdi, 2003; Sacco & Hugenberg, 2009), elicit trait inferences that correspond to the expressions themselves (Adams, Garrido, Albohn, Hess, & Kleck, 2016; Adams et al., 2012; Carré, Morrissey, Mondloch, & McCormick, 2010; Montepare & Dobish, 2003; Oosterhof & Todorov, 2008, 2009; Said et al., 2009; Zebrowitz et al., 2010). For example, faces with downturned eyebrows, associated with angry expressions, are judged to be more dominant (Keating et al., 1981) and less trustworthy (Oosterhof & Todorov, 2008). Similar results have been found for static morphological features: facial width-to-height ratio, linked to testosterone, is associated with perceptions of aggression, physical strength, and dominance (Carré, McCormick, & Mondloch, 2009; Carré et al., 2010; Hehman, Leitner, et al., 2015; Hehman, Leitner, & Gaertner, 2013), and facial height is linked to perceptions of leadership ability (Re, Hunter, et al., 2013; Re, DeBruine, et al., 2013).

Overall, these findings provide examples of factors that drive trait impressions from faces. Many impressions reliably correlate with certain facial features, and some impressions are more substantive than others – for example, traits such as warm/cold, trustworthy, competent, and dominant/weak frequently appear in the literature, but less prominent are other adjectives

that we might use to describe someone, such as creative or fabulous. However, these models do not paint a full picture of face perception (Todorov et al., 2015). Specifically, these models are limited by their reliance on the traditional hypothesis-driven approach, which cannot capture all of the possible impressions that arise: researchers choose (a priori) to examine specific traits but cannot determine which ones are spontaneously evoked during face perception, nor which ones are most critical for driving specific impressions, nor the degree to which clusters of traits (e.g., warm, trustworthy) tend to co-occur (Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Moreover, it is unclear which facial cues (of which there are millions of combinations) are most important for driving trait impressions, and how researchers should define these cues (e.g., mouth versus lips versus image pixel). Finally, these models do not completely account for biases in downstream societal outcomes (Todorov et al., 2013, 2015) – for instance, why outcomes associated with particular trait impressions are moderated by target and perceiver characteristics such as race and gender.

Data-driven methods. To address these limitations, recent research has turned to data-driven methods to model social perceptions of faces These computational approaches are not constrained by prior hypotheses, allow complete control over face stimuli, and focus on explaining the variance in facial appearance that leads to trait impressions such as trustworthiness and competence (for a review, see Todorov et al., 2015). Subsequent work in this area has relied on the extensively-validated statistical face space (Blanz & Vetter, 1999, 2003; Oosterhof & Todorov, 2008; Todorov et al., 2013), a multidimensional space of all possible faces constructed from 3D scans of actual faces, which provides a controlled measure of the variations in features that can be linked to specific traits. These techniques make it possible to model the underlying variations that contribute to specific trait impressions by precisely

manipulating facial features, such as exaggerating the curvature of the mouth from smiling to frowning (Oosterhof & Todorov, 2008). For example, faces evaluated on trustworthiness seem to change from expressing anger to expressing happiness when moving from the negative (least trustworthy) to positive (most trustworthy) extremes (Oosterhof & Todorov, 2008). Similarly, faces evaluated on threat seem to change from appearing feminine and babyfaced to masculine and mature when moving from less threatening (i.e., lower ratings of threat) to more threatening (Oosterhof & Todorov, 2008; Todorov et al., 2013).

Data-driven methods also allow us to control for shared variance between different models and examine whether these models diverge. For instance, models of threatening faces converge with models of dominant faces (Todorov et al., 2015). As models become more similar, it becomes harder to differentiate between faces that vary along each model (Todorov et al., 2013). Consequently, impressions of faces that vary along one trait (e.g., threat) elicit similar impressions on the other trait (e.g., dominance). Critically, this finding corroborates other lines of research in the social cognition literature, that many trait impressions are highly correlated (Fiske et al., 2007; Oosterhof & Todorov, 2008) – indicating that people evaluate faces on a small number of basic, higher-order dimensions. We now consider research on these dimensions.

Major Dimensions Underlying Person Perception

Beyond the limited set of traits discussed in this review, there are countless different ways to evaluate faces and individuals. However, many of these trait impressions – such as competent and intelligent – are highly correlated with one another (Oosterhof & Todorov, 2008; Todorov, Said, et al., 2008). Consequently, past researchers have used data-reduction methods to distill these trait impressions down to the latent theoretical dimensions underlying person perception (Freedman, Leary, Ossario, & Coffey, 1953; Leary, 1957). Research across different

domains (e.g., face perception, group perception) consistently finds at least two dimensions that explain most of the variance in how people evaluate each other (Fiske et al., 2007, 2002; Oosterhof & Todorov, 2008; Sutherland et al., 2013). Furthermore, the theoretical underpinnings of these two-dimensional models are similar across different domains, and emphasize the functional significance of social impressions for survival (i.e., threat appraisal and approach/avoidance behaviour). Thus, facial impressions on these dimensions may reflect evolutionary pressures to quickly identify (1) whether a target's intentions toward us are good or ill (Fiske et al., 2007; Oosterhof & Todorov, 2008), and (2) the target's ability to enact those intentions (Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013). Whereas the first dimension captures traits associated with perceived intent, such as warm, friendly, caring, aggressive, and trustworthy, the second dimension, associated with perceived ability, includes traits such as competent, strong, intelligent, industrious, and skillful (Fiske et al., 2007).

These two dimensions have emerged across independent lines of research, resulting in a somewhat scattered terminology: Trustworthiness and Dominance in the facial first impressions literature (Oosterhof & Todorov, 2008; Todorov, Said, et al., 2008), and Warmth and Competence in the social cognition literature (i.e., abstract beliefs about people and groups; Fiske et al., 2007, 2002). These dimensions have an extensive history dating back to the pivotal work of Solomon Asch (1946).

In Asch's (1946) experiment, evaluations of people described on a number of traits (e.g., determined, practical, industrious, skillful) were unduly influenced by the inclusion of a valence 'warm—cold' trait, demonstrating that Gestalt principles of clustering could be applied to social impressions – a person described as warm and intelligent is wise, but a person described as cold and intelligent is sly (Fiske et al., 2007). Subsequent work on trait clusters revealed that people

tended to categorize traits belonging to the same person along two latent dimensions, social good—bad and intellectual good—bad (Rosenberg, Nelson, & Vivekananthan, 1968). Critically, these two dimensions are nearly orthogonal: ostensibly related traits (e.g., scientific, determined, intelligent, imaginative) cluster around the same region of the two-dimensional space, indicating that evaluations along both dimensions jointly guide abstract impressions of a target. For instance, the dramatic influence of warm—cold in Asch's experiment may simply reflect variance on the social (warmth) dimension while keeping the intellectual dimension constant (Fiske et al., 2007).

While influential, these early behavioural studies were limited in several ways, lacking ecological validity (real-world perception does not always entail reading a list of traits about a person) and being constrained by researcher-selected traits. Research on the two dimensions has proliferated in recent decades, with evidence from face perception and intergroup perception demonstrating that people everywhere may evaluate each other according to their intentions (warmth, trustworthiness) and ability (competence, dominance).

Warmth/Trustworthiness and Competence/Dominance

Dimensions of social cognition. Within person perception research, traits loading on dimensions of warmth (cf. morality; Fiske et al., 2007) and competence were found to account for most of the variance in our perceptions of social behaviour (Wojciszke, Bazinska, & Jaworski, 1998), interpretations of personal events (Wojciszke, 1994), and impressions of prominent people (Wojciszke, 2005). Evaluations of warmth, which encompass traits that overlap with evaluations of morality and trustworthiness (e.g., honest, helpful; Wojciszke, 1994; Wojciszke et al., 1998), involve inferences about the intentions of a target (Fiske et al., 2007; Reeder, Kumar, Hesson-McInnis, & Trafimow, 2002; Zebrowitz & Montepare, 2008), and

inform subsequent decisions to approach or avoid a target (Cacioppo, Gardner, & Berntson, 1997; Peeters, 2002; Slepian, Young, Rule, Weisbuch, & Ambady, 2012). Evaluations of competence, in contrast, are diagnostic of a target's perceived ability to enact their (good or ill) intentions (e.g., intelligent, competent, strong; Fiske et al., 2007; Wojciszke, 1994, 2005). Similar models have emerged in intergroup perception research. In Fiske and colleagues' stereotype content model, warmth and competence together comprise threat appraisal of social groups: warmth reflects the perceived intent of a particular group, whereas competence reflects that group's perceived ability to carry out their intentions (Cuddy et al., 2009; Cuddy, Fiske, & Glick, 2008; Fiske et al., 2007, 2002).

Dimensions extracted from faces. Within the face perception literature, Oosterhof and Todorov (2008) developed a two-dimensional model of trustworthiness and dominance, which jointly guide appraisals of threat. First, they used principle components analysis (PCA) to identify the dimensional structure of traits spontaneously inferred from emotionally-neutral faces. By manipulating features of computer-generated faces, they examined whether evaluations along the two dimensions were sensitive to different facial cues (i.e., whether they diverge). Two dimensions emerged: evaluations of trustworthiness were closest in space to the first latent factor, whereas evaluations of dominance were closest to the second factor. In addition, after removing these two traits from the overall analysis and submitting the remaining traits to a subsequent PCA (controlling for frequency of trait use), they found trustworthiness impressions to be highly correlated with the first (>.90) but not the second dimension (~-.10), whereas dominance impressions were moderately correlated with the second (.53-.77) but not the first dimension (~-.24). Thus, the two-dimensional structure appears to be orthogonal, and trait judgements of trustworthiness and (to a lesser extent) dominance can approximate the underlying

dimensions. These dimensions were subsequently replicated in studies using morphed and averaged real stimuli (Sutherland et al., 2013; Todorov et al., 2013) and reverse correlation (Dotsch & Todorov, 2012; Robinson, Blais, Duncan, Forget, & Fiset, 2014).

Furthermore, evaluations of 'threat' emerged as a function of this two-dimensional model, demonstrating the functional significance of social impressions on these dimensions. Faces independently judged as threatening were both untrustworthy, signalling ill intent, and dominant, signalling the capacity for harm (Oosterhof & Todorov, 2008; Sutherland, Oldmeadow, & Young, 2016; Todorov et al., 2013). An independent model of threat evaluation constructed from vectors of the two-dimensional model was nearly identical to the model based on participants' actual threat judgements. Thus, these findings indicate that models of facial impressions closely agree with underlying theory from person and group perception research (Cuddy et al., 2008; Fiske et al., 2002), that these dimensions hold adaptive value for threat appraisal and survival.

Critically, although there are many reasons to believe that models of warmth and competence from intergroup perception research (Cuddy et al., 2008; Fiske et al., 2007, 2002) converge with models of trustworthiness and dominance from the facial first impressions literature (Oosterhof & Todorov, 2008; Sutherland et al., 2013), recent work by Sutherland and colleagues (2016) suggests that these models are not entirely equivalent. Specifically, whereas trait impressions on dimensions of warmth and trustworthiness (signalling intent) are highly related, impressions of competence and dominance (signalling ability) are less strongly related. Because these two literatures use different stimuli (e.g., faces versus abstract descriptions of individuals and groups), it is possible that indicators of a target's ability may vary across different contexts.

For instance, facial expressions, which signal a target's intentions and emotional state (Adams et al., 2012; Bruce & Young, 1986; Darwin, 1872; Ekman & Friesen, 1971; Zebrowitz & McDonald, 1991), correspond to a greater extent with perceptions on the warmth/trustworthiness dimensions (Oosterhof & Todorov, 2008; Todorov, Baron, et al., 2008; Todorov, Said, et al., 2008). Consistent with both social and facial models of person perception, happy-looking faces signal approachability, whereas angry-looking faces signal ill intent and should be avoided (Fiske et al., 2007, 2002; Oosterhof & Todorov, 2008; Sutherland, Oldmeadow, et al., 2016). Thus, warmth and trustworthiness are strongly related.

In contrast, facial structural cues signal physical ability (i.e., dominance) rather than social or intellectual ability (i.e., competence), as would be the case with abstract cues such as group labels (e.g., Black professional; Fiske et al., 2002). Models of facial dominance judgements correlate with structural facial features such as masculinity (associated with physical strength; Oosterhof & Todorov, 2008; Sutherland, Young, Mootz, & Oldmeadow, 2015), higher facial width-to-height ratio (associated with physical strength and dominance; Hehman, Leitner, et al., 2015; Hehman et al., 2013), and to a lesser extent with higher face height (associated with competence and leadership ability; Re, Hunter, et al., 2013; Re, DeBruine, et al., 2013). Thus, the facial dominance dimension that signals physical ability may be less related to the competence dimension that signals ability through status or prestige (Cuddy et al., 2008; Fiske et al., 2007, 2002; Sutherland, Oldmeadow, et al., 2016). Consistent with this explanation, judgements of competence and dominance were even less related for female faces – perceivers may be less accurate at identifying dominance cues in female faces, and dominance may be less important for evaluating the ability of female targets (Sutherland, Oldmeadow, et al., 2016).

These findings suggest that the importance of competence versus dominance for judgements of ability may vary across different contexts. For instance, facial competence predicts electoral success in politics (Ballew & Todorov, 2007; Chen et al., 2014) whereas facial dominance predicts success in the military (Mueller & Mazur, 1996). Competence and dominance may reflect distinct pathways to achieving power and status (Fiske et al., 2002; Sutherland, Oldmeadow, et al., 2016), and can be subsumed under a broader dimension of "Ability" (Sutherland, Oldmeadow, et al., 2016). Although competence seems to better reflect this ability dimension in certain contexts, with impressions of competence contributing most to this factor (Sutherland, Oldmeadow, et al., 2016) and people more often spontaneously describing faces as competent or intelligent than dominant (Oosterhof & Todorov, 2008; Sutherland et al., 2015), the relative importance of competence versus dominance varies with context – and "ability" incorporates both competence and dominance as different routes to judging a target's ability to help or harm. To avoid confusion, we use the terms Trustworthiness and Dominance moving forward, as these are the terms used most commonly in the face perception literature.

Youthful-attractiveness dimension. In addition, although the two-dimensional model derived from facial impressions appears to be orthogonal, others have found that actual evaluations of individuals on these dimensions are often positively (if moderately) correlated (Fiske et al., 2007; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Rosenberg et al., 1968). Consistent with these findings, more recent work incorporating diverse and naturalistic facial stimuli with a broader age range extracted a third additional dimension of youthful-attractiveness (Sutherland et al., 2013). The emergence of this additional dimension converges with findings from babyfacedness research – impressions associated with babyfaced features load on both

trustworthy (e.g., warm, honest) and dominant (e.g., physically weak) dimensions (Zebrowitz & Montepare, 2008) – as well as research on the well-known halo effect, which finds that attractiveness correlates with judgements on a number of other traits loading on both dimensions (e.g., sociable, competent) when a two-dimensional structure is imposed (Judd et al., 2005; Rosenberg et al., 1968).

Generalizability of the Two-dimensional Model

Across a substantial body of work, these two-dimensional models of trustworthiness (cf. warmth) and dominance (cf. competence) are theoretically presumed to be universal, as noted above, suggesting that humans perceive and evaluate all other humans along both dimensions (Cuddy et al., 2008; Fiske et al., 2007; Judd et al., 2005; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016; Todorov et al., 2013; Todorov, Said, et al., 2008). However, given the manifold complexities of social interaction, do people always use both dimensions to inform impressions of faces (and other people)? There are several reasons to question the presumed independence of these two dimensions. First, when people evaluate individuals in abstract terms, their impressions on these two dimensions are often found to correlate positively (if moderately), consistent with the halo effect (Fiske et al., 2007; Judd et al., 2005; Rosenberg et al., 1968). However, when people evaluate social *outgroups*, their impressions along the two dimensions are often correlated: certain groups are judged as higher on one dimension and lower on the other in stereotypically-consistent ways (e.g., female professionals as low on warmth but high on ability, the elderly as high on warmth but low on ability; for a review, see Fiske et al., 2007). Evidently, the two dimensions of person perception are not always independent.

Critically, although models derived from the facial first impressions literature find these two dimensions to be orthogonal, these models were constructed from White participants evaluating White male targets (Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov et al., 2013). The type of facial stimuli (e.g., real vs. computer-generated, diverse vs. homogenous, controlled vs. naturalistic) used in these models can considerably influence which trait impressions spontaneously arise (Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland, Oldmeadow, et al., 2016), how much variance exists around these trait impressions (Burton, Kramer, Ritchie, & Jenkins, 2015; Hehman, Sutherland, Flake, & Slepian, 2017; Jenkins, White, Van Montfort, & Burton, 2011; Sutherland et al., 2013), and even the number of dimensions extracted (Sutherland et al., 2013; Todorov et al., 2015). If the two dimensions of face perception are indeed correlated for cross-group impressions, as they are in models of group perception (Fiske et al., 2002, 2007), it is critical to test these models with a rigorous measurement framework capable of capturing variance on these social categories (e.g., race, gender). Yet to date, the generalizability of these prominent models across social categories has not been scrutinized.

Finally, the orthogonal two-dimensional model is insufficient to explain systematic biases in downstream societal outcomes of facial impressions (Todorov et al., 2015). The impact of facial impressions along these dimensions vary with target characteristics such as gender and race, often in systematic ways (Blair et al., 2004; Carpinella et al., 2015; Eberhardt et al., 2006; Pillemer et al., 2014; Rule & Ambady, 2009; Sutherland et al., 2015). For instance, whereas facial competence predicts positive outcomes for male CEOs, this effect is diminished for female CEOs (Pillemer et al., 2014) and inverted for Black CEOs (Livingston & Pearce, 2009); instead,

facial warmth predicted positive outcomes for both groups, indicating that warmth may be more important than competence in driving positive evaluations of female and Black CEOs.

Overall, there is reason to believe that the orthogonal two-dimensional model may not generalize to other social groups beyond the in-group. Social category knowledge has profound effects on face perception – thus, it is important to consider research on group biases.

Intergroup Effects in Face Perception

Faces signal not only the intentions and ability of a target, but also their social identity (Cohen, 1981; Freeman & Ambady, 2011; Freeman et al., 2010; Hehman, Ingbretsen, et al., 2014; Kubota & Ito, 2007; Macrae & Bodenhausen, 2000; Tajfel, Billig, Bundy, & Flament, 1971). For instance, race, gender, typicality, and even sexual orientation are readily identifiable from faces (Adolphs et al., 2016; Bruce & Young, 1986; Calder & Young, 2005; Freeman et al., 2010; Haxby et al., 2000; Hehman, Ingbretsen, et al., 2014; Kubota & Ito, 2007; Oosterhof & Todorov, 2008; Rule & Ambady, 2008a; Rule, Ambady, & Hallett, 2009). These social categories are highly salient because people readily categorize themselves and others along these categories (Cohen, 1981; Freeman & Ambady, 2011; Tajfel et al., 1971), and generalize group-level attributions (e.g., the perceived intent and ability of a particular group) to attributions about individuals who belong to these groups in order to streamline the perceptual process (Cuddy et al., 2008; Fiske et al., 2002, 2007; Hehman, Ingbretsen, et al., 2014; Macrae & Bodenhausen, 2000; Tajfel et al., 1971). Furthermore, people also categorize themselves as members of social groups (Tajfel & Turner, 1986; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Within social cognition, a substantial body of work has continuously demonstrated the pervasive influence of intergroup bias: we reliably perceive the world in terms of "us" (ingroups) and "them" (outgroups), favouring those who belong to ingroups and expressing prejudice against outgroups

even under the most minimal group conditions (Cikara & Van Bavel, 2014; Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993; Macrae & Bodenhausen, 2000; Ratner & Amodio, 2013; Tajfel et al., 1971; Tajfel & Turner, 1986; Turner et al., 1987). Critically, these intergroup biases manifest as more positive impressions and representations of ingroup versus outgroup members; even minimal ingroup faces are perceived as warmer and more trustworthy than outgroup faces (Gaertner et al., 1993; Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014; Todorov et al., 2015).

Social-cognitive theories about the mechanisms underlying intergroup biases emphasize the category salience of individuals (Freeman & Ambady, 2011; Kubota & Ito, 2007; Neuberg & Fiske, 1987; Tajfel et al., 1971), suggesting that merely categorizing people into groups elicits perceptual biases, altering the type or depth of processing that ingroup and outgroup targets receive (Bernstein, Young, & Hugenberg, 2007; Cikara & Van Bavel, 2014; Cohen, 1981; Freeman, Stolier, Ingbretsen, & Hehman, 2014; Hehman, Carpinella, et al., 2014; Hehman et al., 2010; Hugenberg, Miller, & Claypool, 2007; Hugenberg, Young, Bernstein, & Sacco, 2010; Ratner & Amodio, 2013). Specifically, perceivers tend to process outgroup faces more categorically – using social category information to inform impressions about the individual (Hugenberg & Sacco, 2008; Hugenberg et al., 2010; Ratner et al., 2014; Tajfel et al., 1971; Young & Hugenberg, 2010) – and ingroup faces in a more individuated manner. These theories are anchored in a "cognitive-miser" (Freeman & Ambady, 2011; Martin & Macrae, 2007) view of person perception, which proposes that individuated perception requires more cognitive resources than categorization. Thus, because social categorization allows perceivers to process targets efficiently, cross-group perceptions are subject to motivational influences, stereotype activation, and prejudice (Bargh, 1999; Brewer, 1988; Cuddy et al., 2008; Dovidio, Kawakami,

Johnson, Johnson, & Howard, 1997; Fiske et al., 2007, 2002; Fiske & Neuberg, 1990; Freeman & Ambady, 2011; Freeman et al., 2010; Hugenberg & Bodenhausen, 2003; Hugenberg & Sacco, 2008; Sinclair & Kunda, 1999).

Given the pervasiveness of intergroup biases in face perception, I propose that the two major dimensions underlying facial impressions – trustworthiness (cf. warmth) and dominance (cf. competence) – may be correlated across group boundaries. First, consider that if ingroup faces are perceived in a more individuated manner, then perceivers use facial cues to inform their evaluations of a target's intentions and ability (Freeman & Ambady, 2011; Neuberg & Fiske, 1987; Zebrowitz et al., 2003; Zebrowitz & Montepare, 2008), thus allowing evaluations of a target to be jointly guided by two orthogonal dimensions. However, outgroup faces, which are processed more categorically, are subject to social categorical attributions – impressions along one dimension may bias the other in systematic ways depending on the social identities that are salient, as well as the stereotypes associated with those categories (stereotype content model; Cuddy et al., 2008; Fiske et al., 2007, 2002). Although no study to date has tested a model in which the two-dimensional model is allowed to correlate for cross-group impressions, there is a wealth of evidence that indicates these dimensions may be correlated when perceiving outgroup faces.

Activation of stereotype-congruent impressions. The impact of stereotypes on perception is pervasive, and extends simultaneously to societal ingroups – which are favoured and evaluated highly on both warmth and competence – as well as societal outgroups, which tend to be evaluated ambivalently in stereotype-congruent ways (Cuddy et al., 2008; Fiske et al., 2007, 2002). In the US, groups perceived as low-warmth and high-competence include rich people, Asians, and female and minority professionals, and groups perceived as high-warmth and

low-competence include the elderly and the physically or mentally disabled (Fiske et al., 2007, 2002). Critically, perceptions of outgroup members are influenced in stereotype-congruent ways.

For instance, research dating to the work of Secord and Bevan (1956) finds that perceivers apply stereotypes to faces classified as Black no matter how race-typical these faces look, and subsequent impressions (e.g., aggressive) and judgements (e.g., likability) are influenced by the activation of the Black category (Bargh, 1999; Brewer, 1988; Devine, 1989; Dovidio et al., 1997; Fazio, Jackson, Dunton, & Williams, 1995; Fiske & Neuberg, 1990; Hehman, Volpert, & Simons, 2013; Sinclair & Kunda, 1999). Among perceivers both high and low in prejudice, stereotype-congruent impressions (e.g., Blacks and violence) were automatically activated when presented with a stereotyped outgroup actor (Devine, 1989). For example, videos of ambiguously aggressive behaviours, such as bumping in the hallway, were perceived as meaner and more threatening (i.e., more ill-intentioned and having greater capacity for harm) when the target was a Black actor compared to a White actor (Devine, 1989), suggesting that evaluations of threat were more accessible. In a similar vein, traditional women (i.e., homemakers) are perceived as higher in warmth but less competent compared to professional women, whereas professional women are perceived as low in warmth but more competent (Cuddy, Fiske, & Glick, 2004; Fiske et al., 2002; Fiske, Xu, Cuddy, & Glick, 1999). Furthermore, when professional women become mothers, they trade perceived competence for perceived warmth, whereas professional men maintain perceived competence while gaining perceived warmth (Cuddy et al., 2004) – suggesting that the dimensions of warmth and competence may be correlated negatively for impressions of outgroup targets.

Similarly, perceivers evaluate outgroup targets differently when presented with information about the targets' apparent intentions (Sinclair & Kunda, 1999, 2000). For instance,

perceivers who received either positive or negative feedback from a Black professional, constituting the manipulation of the outgroup member's perceived intent (i.e., being praised generally signals good intentions), rated the target as more competent when they had received positive feedback, and less competent when they had received negative feedback (Sinclair & Kunda, 1999). Similarly, female (but not male) university instructors were perceived as less competent by students who had received negative feedback from them (Sinclair & Kunda, 2000). These findings suggest that impressions of a target's intentions influence subsequent impressions of ability in stereotype-congruent ways. Therefore, the dimensions of warmth and competence may be related for cross-group impressions.

Contextual influences on perceptions of threat. A broad swathe of research by Fessler and colleagues also illustrates the influence of context in perceptions of threat: whether or not a target looks particularly threatening depends on perceiver characteristics, such as parenthood (Fessler, Holbrook, Pollack, & Hahn-Holbrook, 2014), being in a group (Fessler & Holbrook, 2013b), being incapacitated (Fessler & Holbrook, 2013a), as well target characteristics, such as race and perceived competence (Holbrook et al., 2016). For instance, while individuals perceived as physically threatening (i.e., dominant) are perceived to be physically larger (Holbrook et al., 2016), this effect also extends to high-status (i.e., competent) individuals if they belong to stereotypically safe racial groups (e.g., White; Holbrook et al., 2016). Moreover, physical size (signalling capacity for harm) mediates group-based inferences of threat (Holbrook et al., 2016).

In another study, physically restrained participants – compared to participants who were unrestrained – perceived an angry-looking man (signalling ill intent) as being physically larger and more muscular (signalling capacity for harm), indicating that threat appraisals may drive correlations between perceptions of intentions and perceptions of ability (Fessler & Holbrook,

2013a). In this case, being perceived as having negative intentions alone subsequently predicted perceptions of high ability. In contrast, however, greater perceived ability might also lead to impressions of ill intent: targets in another study were perceived as angrier-looking when they were holding household objects that could be used as weapons (e.g., gardening shears versus a watering can; Holbrook et al., 2014). In this case, being perceived as high in ability alone subsequently predicted perceptions of low intentions. In a complex social world, it is more adaptive to minimize risk. Thus, the greater the capacity for harm, the costlier it is to underestimate the (potentially ill) intentions of a target: illustrating various contexts where it would be more adaptive for evaluations of ability to negatively correlate with evaluations of intent.

These findings serve as compelling examples of the pervasiveness of intergroup biases in person perception, as well as various situations in which cross-group impressions on the dimensions of warmth and competence can be correlated. Theories of threat management attribute these cross-group effects to the functional significance of threat appraisal. Since impressions of warmth and competence arise within milliseconds, we may not always have the time or opportunity to assess outgroup members on both dimensions. Evolutionary theories of threat management stipulate that it is better to be safe than sorry (Bar et al., 2006; Fessler et al., 2014; Fiske et al., 2007; Haselton & Nettle, 2006; Holbrook et al., 2014), to exercise caution when interacting with targets perceived as being low in warmth *or* high in competence. These first impressions arise within milliseconds of viewing a target – thus, the most salient aspects of their appearance, such as their social identity, may unduly influence our evaluations of their warmth and competence. For these reasons, it may be more adaptive to assume the worst of targets who fulfill at least one of these (potentially threatening) conditions, and avoid them

accordingly. Consistent with this framework, extant research on the Compensation (Yzerbyt, Provost, & Corneille, 2005) and Innuendo effect (Kervyn, Bergsieker, & Fiske, 2012) finds that when a target is high on one dimension, they are subsequently inferred to be low on the other. Indeed, recent work in the facial first impressions has replicated the independence of these two dimensions for impressions of White male targets, but not of female targets – consistent with previous findings, those perceived as more trustworthy were also perceived as less dominant (Sutherland et al., 2015).

Alternative explanations for the fluidity of dimensions in cross-group impressions include (1) theories of instrumentality or functional utility, in which perceivers may evaluate targets on a greater number of dimensions if the target is perceived as highly instrumental to the perceiver, capable of facilitating the perceiver (Esses, Jackson, Dovidio, & Hodson, 2005; Fiske et al., 2007; Guerra, Gaertner, António, & Deegan, 2015); (2) familiarity or social closeness, in which novel faces that resemble the faces of perceivers or their significant others are evaluated similarly to the self or significant other (Günaydin, Zayas, Selcuk, & Hazan, 2012; Verosky & Todorov, 2010a, 2010b, 2013), and (3) overarching theories of perceptual expertise, which suggest that our ability to individuate faces improves with practice – we may simply be less experienced at viewing outgroup faces (Bukach, Gauthier, & Tarr, 2006; Valentine, 1991; Young, Hugenberg, Bernstein, & Sacco, 2012). In fact, it's possible that the diagnosticity of the two dimensions are asymmetrical (Fiske et al., 2007), such that one is always more important or diagnostic depending on the context. Research on the primacy of warmth suggests, for example, that evaluations of warmth occur before evaluations of competence (e.g., faces viewed within 100ms are judged mostly reliably on trustworthiness/warmth, followed by competence; Willis & Todorov, 2006) – it may be that in order to evaluate a target on a second dimension, their

intentions (cf. warmth) would first have to be perceived as positive. Since current prominent models of face perception were derived from White perceivers evaluating White male targets, the independence of the two dimensions may be unique to perceptions of ingroup members.

Although, in the current research, I was unable to distinguish between these theoretical accounts, my goal was to first establish whether current two- and three-dimensional models of facial impressions could generalize across group boundaries.

The Present Study

Based on the rationale presented above, the current research tests classic models of face perception, hypothesizing that Trustworthiness and Dominance vary in the extent to which they correlate as a function of race and gender, and are not always independent. Despite extant research indicating that Trustworthiness and Dominance are not independent when forming impressions of outgroup targets, the generalizability of current prominent models of person perception has not been rigorously scrutinized. Critically, previous work in this area has been limited by methodological concerns such as a reliance on principle components analysis (PCA), a lack of broadly representative facial stimuli, and limited sample sizes of ratings. Therefore, building on this theoretical framework, this thesis aims to advance the person perception literature by building and testing an extended model of person perception.

The current research explores this possibility, testing the hypotheses that dimensions of person perception are correlated for perception across group boundaries, both as a function of race and gender. Specifically, I fit the classic two- and three-dimensional models to ratings of faces across trait impressions, testing model fit to assess how well these models explain the variance in observed ratings, and whether the correlation between factors of Trustworthiness (cf.

warmth) and Dominance (cf. competence) vary depending on the race and gender of the target and perceiver.

Evidence in support of my hypothesis, and contrary to existing models, would reveal poorer-fitting models when the two-dimensional factor structure is applied to other-group impressions. That is, I expected two-dimensional models for other-group impressions to have worse model fit compared to own-group impressions (Hypothesis 1). Second, I predicted better model fit for other-group impressions when the two dimensions are allowed to correlate (Hypothesis 2), which suggests that perceptions across groups are more unidimensional than perceptions of one's own group. Furthermore, based on theories of threat management, I predicted that impressions on these dimensions will be correlated negatively (Hypothesis 3). Finally, I expected to replicate the classic two-dimensional structure for own-group perceptions (Hypothesis 4). These results would generalize models of person perception to individuals beyond White males, informing our understanding of how such impressions are formed.

Biases in person perception influence social judgements and reverberate throughout society, and people use these models to predict real-world outcomes (Ballew & Todorov, 2007; Eberhardt et al., 2006). Thus, a model that is able to accurately capture universal dimensions of social cognition would allow researchers to (1) predict downstream societal outcomes more accurately, (2) examine systematic biases across these real-world outcomes, (3) understand why and how cross-race effects occur (e.g., in what direction are the dimensions correlated, and under what circumstances might that change?), (4) explore the circumstances in which perceptions are more likely to be "unidimensional" (e.g., are all cross-group impressions unidimensional?), and (5) understand how to help people differentiate each other on a greater number of dimensions, perceiving each other more multidimensionally.

Method

Analytic Approach

To assess whether the dimensions of person perception are flexible (varying in the extent to which they are correlated) as a function of race and gender, I built two separate grouped factor structure models for *own-group* and *other-group* perceptions. Specifically, I focus on White participants' perceptions of male and female targets from three racial groups, White – to replicate previous models and act as a reference for comparisons – East Asian, and Black. These groups were selected for their perceptual salience, availability of stimuli, and differing stereotype content: race and gender are both strong indicators of social group membership, and extant research finds systematic, stereotype-congruent differences in perceptions of individuals from these outgroup categories (e.g., Black male as less trustworthy; Freeman & Ambady, 2011; Hehman, Ingbretsen, et al., 2014; Sutherland et al., 2015; Todorov, Baron, et al., 2008; Zebrowitz et al., 2010).

Evidence in support of our hypothesis, and contrary to existing models, would reveal (1) poorer model fit when applying the two-dimensional structure to *other-group* compared to *own-group* perceptions, and (2) stronger correlations among the presumably-independent dimensions of warmth and competence for other-group impressions, particularly in stereotype-congruent ways. For example, since other-race individuals are perceived as being more ill-intentioned under a threat management framework (Cuddy et al., 2008; Fiske et al., 2002, 2007), we predict that targets perceived as lower on the warmth dimension will simultaneously be perceived as higher on the competence dimension, and vice versa: person perception across group boundaries will be more unidimensional.

To develop and compare these models, I use a rigorous measurement approach, conducting confirmatory factor analysis (CFA) in a structural equation modeling (SEM) framework. CFA is a confirmatory technique: researchers build and test a theory-driven model (e.g., the classic two-dimensional model of Trustworthiness and Dominance) to estimate a population covariance matrix of traits that I compare with the observed covariance matrix of traits (Schreiber et al., 2006; Ullman & Bentler, 2013). Thus, SEM allows for comparisons between the estimated and observed matrices while accounting for measurement error. The core of SEM analysis involves the examination of the coefficients of hypothesized relationships (i.e., the loadings of each trait on their latent dimensions, and whether Trustworthiness and Dominance are more correlated for other-group perceptions) and should indicate whether the hypothesized model was a good fit to the observed data (i.e., the two-dimensional model should have better model fit for own-group perceptions).

The advantages of SEM compared to PCA are manifold, allowing researchers to (1) estimate the multidimensional structure of social impressions by partitioning the total covariance among indicators (i.e., trait impressions) loading on each latent factor, (2) estimate the relationships between latent factors corrected for measurement error, and (3) systematically compare alternative models of the data based on model fit indices (Jöreskog, 1970; Morin, Arens, & Marsh, 2016; Schreiber et al., 2006; Ullman & Bentler, 2013). PCA is not a good tool if researchers want to make conclusions about generality, as it was never intended to be used to build generalizable models – only to best explain the observed data (for a review, see Field, 2009). Thus, there are some assumptions in PCA that do not allow for generalizability. Although the results of PCA should be similar to the results of factor analysis with a large number of indicators (e.g., more than 30 indicators), the foundational work on current two- and three-

dimensional models of face perception were conducted with less than 30 indicators (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Finally, since PCA is primarily used to explain variance, it does not provide an index of model fit (i.e., how well the solution captures the overall variance in the observed data).

CFA models in SEM are able to do this. SEM therefore make it possible to accurately estimate the underlying dimensions of social impressions for both own-group and other-group impressions, and provides an output containing the different models' goodness-of-fit. I discuss these model fit indices in greater detail in the Results section. All analyses will be conducted using the SEM R package, *lavaan* (Rosseel, 2012).

Representational similarity analysis. In a subsequent analysis, I tested the hypotheses using a different analytic approach in order to address a key limitation of CFA – poor model fit due to its reliance on the highly restrictive independent cluster framework, in which cross loadings between indicators and non-target factors are fixed to zero (seldom the case with real data). For instance, some indicators that are not a priori set to load on the same target factors (e.g., the 'competent' and 'friendly' indicators, which theoretically load on different dimensions) are not allowed to share variance. This usually results in poor model fit because many traits do cross-load in reality, possibly due to the halo effect. I cannot cut these indicators from the model as I would items on a scale, since people do spontaneously evaluate others on traits like attractiveness and competence even if these traits cross-load (Oosterhof & Todorov, 2008; Sutherland, 2015; Sutherland et al., 2013; Todorov et al., 2015). A model using the CFA framework with two latent factors will not be able to capture this effect, because shared variance between indicators across the two latent dimensions are not being captured. This usually results

in poor model fit, which renders the parameter estimates of the model less reliable and interpretable.

Even if these models fit poorly, there may still be interesting patterns in the data that merit exploration. Because the parameter estimates of these models cannot be relied upon, however, as an alternative analytic approach, I also test the models in a different analytical framework, which utilizes the representational similarity analysis (RSA) technique from systems neuroscience (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013; Nili et al., 2014). Originally developed to test computational models of the brain by comparing representations of brain and behavioural data, the RSA technique was recently adapted by Stolier and colleagues (2018) to describe face-based trait perception and the social trait space (Stolier, Hehman, & Freeman, 2018). The advantages of RSA are manifold: to summarize, it allows for the comparison of "trait spaces" (i.e., correlation matrices of traits) across different groups (e.g., own-group vs. other-group perceptions). Critically, RSA captures the *similarity* of trait pairs on the basis of some specific factor, such as the race and gender of the target, which allows me to compare trait spaces across own-group and other-group perceptions.

To elaborate, in an RSA framework, the trait space is conceptualized as a network of trait relationships in which the association between each pair of traits has a quantifiable weight (Kriegeskorte, 2008; Nili et al., 2014; Stolier et al., 2018). For instance, the traits 'warmth' and 'trustworthiness' may be more strongly associated than 'warmth' and 'dominance', giving it greater weight. This trait space, or correlation matrix of traits, can then be represented by a similarity matrix, in which each cell of the matrix is the weighted association (i.e., similarity) of each pair of traits (Diedrichsen & Kriegeskorte, 2017; Stolier et al., 2018).

The RSA technique thus allows researchers to capture the similarity of any given pair of traits (Stolier et al., 2018). Given a set of 14 x 14 trait matrices where each trait (e.g., competence) has a similarity weighting with another trait (e.g., intelligence), each matrix can be compared with another to describe the similarity of trait pairs based on some other factor. Here, that factor would be the race and gender of the target group – when participants are evaluating own-group targets versus other-group targets, how do the similarity of trait pairs differ? RSA allows for the quantitative comparison of these similarity matrices, reflecting a combination of the corresponding similarity in both trait concepts (e.g., warmth and competence) and group concepts (e.g., perceptions of Asian men). Thus, the RSA framework presents an alternative to the CFA, allowing me to explore patterns in trait matrices across different groups if the CFA models cannot be interpreted.

Source of the Data

I fit the classic two-dimensional and three-dimensional factor structures to ratings of faces on 14 trait impressions previously demonstrated to load highly on the major dimensions of warmth, competence, and youthful/attractiveness: aggressive, assertive, attractive, caring, competent, dominant, friendly, healthy, intelligent, smart, physically strong, trustworthy, warm, and youthful (Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov, Said, et al., 2008). Extant research indicates that these traits are spontaneously inferred when perceivers evaluate others (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov, Baron, et al., 2008).

While 5 out of these 14 traits were not included in Oosterhof and Todorov's (2008) original two-dimensional model of face perception (i.e., warm, healthy, youthful, competent, and smart), we included them because they were semantically similar to other traits in their model

(e.g., “intelligent” and “smart”), they were present in subsequent replications of Oosterhof and Todorov’s model (Sutherland et al., 2013; Todorov et al., 2013, 2011), and extant evidence indicates that these traits are indeed highly correlated and spontaneously arise in social impressions (Asch, 1946; Fiske et al., 2007; Rosenberg et al., 1968; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016; Todorov et al., 2015; Zebrowitz & Montepare, 2008). For instance, although “warmth” was not included as a trait in Oosterhof and Todorov’s model, people spontaneously evaluate others on warmth (Asch, 1946; Cuddy et al., 2008; Fiske et al., 2007; Zebrowitz & Montepare, 2008), and warmth correlates strongly with friendliness – a trait that was present in their model (Cuddy et al., 2009; Fiske et al., 2007; Johnson, Freeman, & Pauker, 2012; Nisbett & Wilson, 1977; Sutherland, 2015).

Participants were recruited from MTurk across the US with monetary compensation, and rated real facial stimuli from standardized databases, which offer a tightly-controlled set of stimuli that are broadly representative of facial variation, and are sufficiently diverse to allow our estimates to generalize beyond these samples (See Figure 2). These databases include the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), the MR2 (Strohminger et al., 2016), the Radboud Faces Database (Langner et al., 2010), the NimStim Face Stimulus Set (Tottenham et al., 2009), the Center for Vital Longevity Face Database (Minear & Park, 2004), the OSLO Face Database (Chelnokova et al., 2014), the Eberhardt Face Database (Eberhardt, n.d.), the Face Research Lab London Set (DeBruine & Jones, 2017), and the CUHK Face Sketch database (Wang & Tang, 2009).

Ratings on these 14 trait impressions are made on 1-“Not at all” to 7-“Very much” Likert scales (e.g., “How trustworthy is this person?”). Stimuli targets are presented in random order, and participants rate each target on only one trait such that all ratings are between-subjects – an

approach consistent with previous work in the facial first impressions literature (Hehman, Flake, et al., 2015; Hehman, Leitner, Deegan, & Gaertner, 2013; Oosterhof & Todorov, 2008; Sutherland et al., 2013). For each stimulus target, trait ratings were averaged, such that the mean rating on each of the 14 traits for each stimulus served as the unit of analysis. In our SEM models, these traits were specified to load on their theoretical latent factors, using both the prominent two-factor models and recent three-factor models to test for "unidimensionality" of dimensions across groups.

Figure 2. Example stimuli.

Participants

Based on recommended sample sizes for confirmatory factor analysis in SEM (~150 cases for each indicator for a two-factor model with >6 indicators loading on each factor; Wolf, Harrington, Clark, & Miller, 2013), we collected ratings of faces from 5,040 participants across North America using Amazon Mechanical Turk. Each participant rated 60 or 90 (two phases of data collection) different faces (male or female), of which an equal proportion were Black, East Asian, or White; stimuli were presented to participants in randomized order. Previous work found face ratings to stabilize after 40 participants (Todorov et al., 2005), and additional raters do not significantly alter the average rating. Online participants were drawn from the Mturk Worker pool across the United States and Canada, include workers ages 18-85 with an approval rating above 90%, and received monetary compensation through Mturk. One advantage of MTurk is access to more representative samples compared to undergraduate participant pools (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011).

Data were cleaned in accordance with our pre-registered data cleaning procedure based on response time and frequency of repeated ratings [https://osf.io/65tpb/]. Participants who had relevant demographic information (i.e., race, gender) were included. These criteria resulted in 402,473 ratings of trait impressions across 5,040 participants and 873 stimuli. Participants were aged 17 to 80 (M_{age} = 36.24, SD_{age} = 11.89, 68.1% female), and were 72.6% non-Hispanic White, 10.4% Black, 5.6% Asian, and 11.4% other ethnic minorities which include mixed-race.

Stimuli

In total, stimuli consisted of 299 White (49.8% female), 295 Black (49.2% female), and 279 East Asian (46.2% female) standardized colour photographs of faces from controlled databases. Norming data for the stimuli are available online for certain database websites (e.g., Chicago Face Database, MR2). All stimuli depict frontal views of faces with neutral expressions.

Faces were resized to 611px (wide) x 430px (high), and presented against a plain background. Overall, these stimuli are representative of those used in impression formation research, providing a medley of controlled stimuli. This heterogeneity of stimuli helps to ensure the generalizability of our estimates to other samples (Burton et al., 2015; Jenkins et al., 2011).

Results

Two-Dimensional Models

I fit the classic two-dimensional factor structure to ratings of own-group and other-group faces by partitioning the data according to participant race, building six different models for perceptions of female and male White, Black, and Asian faces. The specific trait loadings on each dimension were the same across all models: trustworthy, aggressive, warm, friendly, caring, attractive, healthy, and youthful formed the Trustworthiness dimension, whereas dominant, competent, physically strong, assertive, intelligent, and smart formed the Dominance dimension (Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Zebrowitz & Montepare, 2008). See Figure 3 for a simple path diagram of this two-factor model.

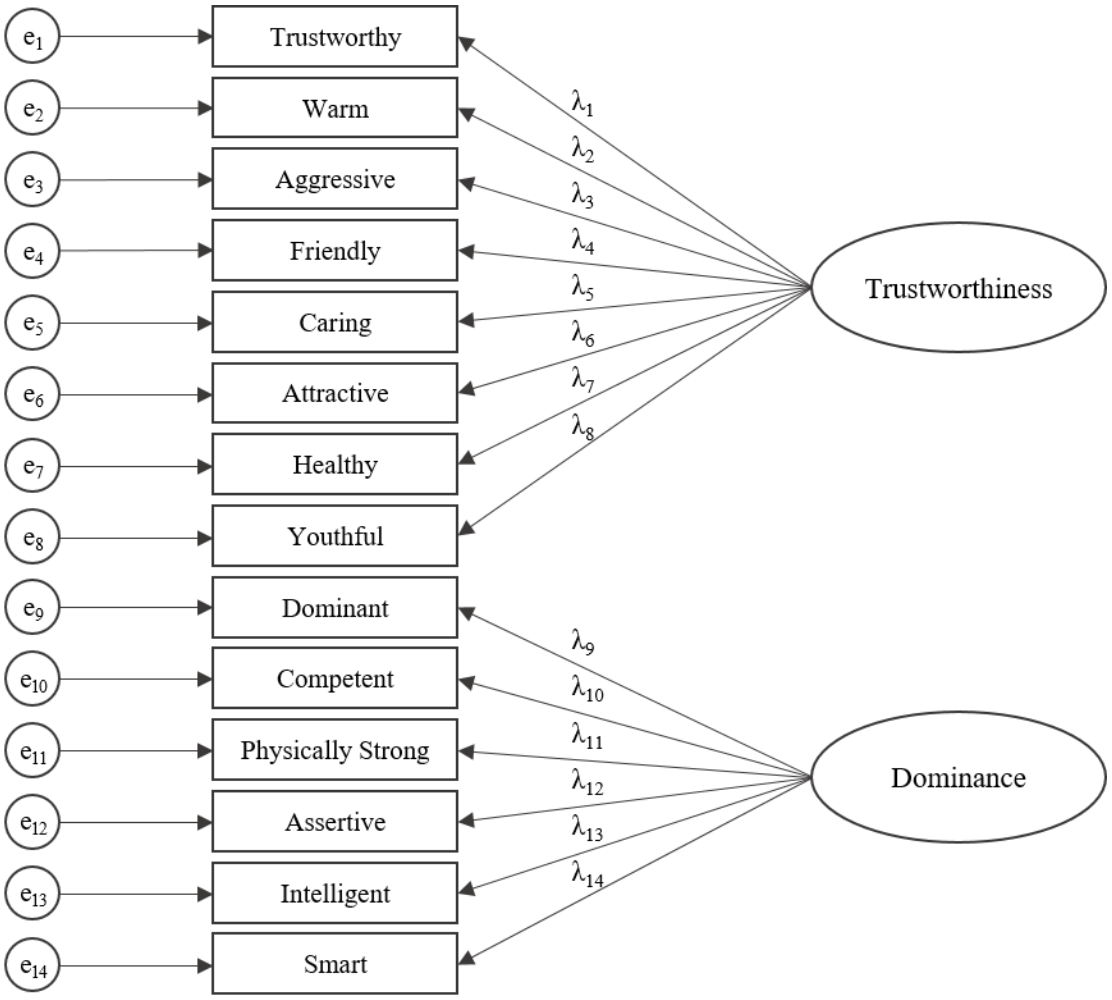


Figure 3. SEM path diagram for a confirmatory factor analysis, representing the loadings of 14 traits onto their latent dimensions of Trustworthiness and Dominance, which were fixed to be orthogonal. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait (e.g., trustworthy) served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable.

Figure 3 shows our first model. This hypothesized model was tested to determine how well classic models of face perception fit our observed data. The latent factors are Trustworthiness and Dominance, and the observed variables, or indicators, are the traits. For each trait, the unit of analysis is the mean rating of a stimulus (across all participants). Trait

indicators have the error term e_i, and λ_i is the regression coefficient indicating the loadings of each trait onto their latent dimension. The two dimensions are fixed to be independent, and cross-loadings between indicators and non-target factors are fixed to zero.

These trait loadings were informed by previous research in face and person perception. As noted earlier, although 5 out of these 14 traits were not included in Oosterhof and Todorov's (2008) original two-dimensional model of face perception (i.e., warm, healthy, youthful, competent, and smart), they were included in the current work due to their theoretical relevance. These traits were present in replications of Oosterhof and Todorov's model (Sutherland et al., 2013; Todorov et al., 2013, 2011), and past research indicates that these traits are highly correlated and arise spontaneously in facial impressions (Asch, 1946; Fiske et al., 2007; Rosenberg et al., 1968; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016; Todorov et al., 2015; Zebrowitz & Montepare, 2008).

Model fit. To assess goodness-of-fit, we followed common guidelines in the literature (for a review, see Schreiber et al., 2006; Ullman & Bentler, 2013). Researchers recommend several goodness-of-fit indicators, including the Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and chi-squared analysis (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996; Yu, 2002). For continuous data, Hu and Bentler (1999) suggest the following cut-offs of acceptable fit: CFI > .95, RMSEA < .06 to .08, and SRMR < .08. The chi-squared statistic is the most widely used summary statistic for assessing model fit, and is primarily used in model modification (i.e., to build better fitting or more parsimonious models), providing a test of whether the modified model is statistically superior to the original. Because the chi-squared test is conservative, it is usually supplemented with the aforementioned fit indices (CFI, RMSEA, and SRMR). Although

the present study does not focus on model modification, we include chi-square values in line with common reporting practices.

We compared model fit indices across various own-group and cross-group perceptions, expecting the two-factor structure to produce better model fit for own-group perceptions (i.e., White participants evaluating White male or female targets) compared to cross-group perceptions. Contrary to my hypothesis, however, results revealed extremely poor model fit across all 6 models – indicating that the classic two-dimensional structure may not reflect how people actually evaluate own-group *or* other-group targets. Critically, by conducting a CFA in an SEM framework, I was unable to replicate Todorov and Oosterhof's (2008) two-dimensional structure of Trustworthiness and Dominance. See Table 1 for a full report of model fit indices. I discuss the implications of this finding in the General Discussion.

Table 1

Indices of model fit for the two-dimensional model of face perception across own-group and cross-group perceptions.

	Chi-square ($\chi 2$)	*p* ($\chi 2$)	CFI	RMSEA	SRMR
Model					
White perceivers rating White males	1096.43	<.001	0.507	0.299	0.330
White perceivers rating Black males	923.79	<.001	0.518	0.273	0.294
White perceivers rating Asian males	675.86	<.001	0.602	0.229	0.171
White perceivers rating White females	1142.09	<.001	0.492	0.307	0.360
White perceivers rating Black females	749.389	<.001	0.700	0.247	0.172
White perceivers rating Asian females	769.174	<.001	0.559	0.266	0.289

Note. CFI = Comparative fit index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean squared residual.

Below are general rules for acceptable fit if data are continuous (Hu & Bentler, 1999; Schreiber et al., 2006):

1. Chi-square: Ratio of $\chi 2$ to $df \leq 2$ *or* 3.
2. CFI: $\geq .95$ for acceptance.
3. RMSEA: < .06 to .08.
4. SRMR: $\leq .08$

These model fit indexes fall well below common cut-offs for acceptable fit. For instance, CFI should be > .90 or .95 for acceptance, but the highest CFI across all models (for White perceivers rating Black women) was only .70. Critically, even the own-group samples failed to produce good model fit, indicating that we were unable to confirm Todorov and colleagues' (2008, 2011, 2013) classic two-factor exploratory structure with our confirmatory analyses.

Formally, the next step to take when a hypothesized model fits poorly is to examine parameter estimates and conduct model modifications to the original model, in order to produce a better fitting model (Schreiber et al., 2006; Ullman & Bentler, 2013). These modifications must be confirmatory in nature, such that they make theoretical sense – as such, there are restrictions in the type of modifications that can be made. Otherwise, it becomes an exploratory process. In the present research, it does not make sense to modify the model by removing traits that fit poorly and keeping others that fit well, because people do spontaneously evaluate others on all of these traits even if the traits do not load well onto the two-dimensional model. As such, the only modification I made was to ensure that the model matched the classic two-dimensional model (Oosterhof & Todorov, 2008; Todorov et al., 2013, 2011) as closely as possible, by removing the five traits that were not present in the original model: warm, healthy, youthful, competent, and smart. Even after removing these traits, model fit did not improve – all model fit indices across all 6 groups still fell well below their acceptable ranges. Since model fit did not improve after removing the five extra traits, I focus on the two-dimensional model with all 14 traits included in subsequent sections.

The parameter estimates of the models (i.e. with all 14 traits) are reported in Table 2. However, I caution against over-interpreting these estimates – if the model does not fit, then the parameter estimates are potentially unreliable (Schreiber et al., 2006; Ullman & Bentler, 2013).

These factor loadings can be interpreted like regression coefficients – for example, for each unit increase in the latent "Trustworthiness" score, the model predicts a *B*-unit increase in "warmth". In addition, the output provides an estimate of covariance between the two latent factors of Trustworthiness and Dominance. If this covariance is significant, then the two dimensions are *not* independent as presumed.

Table 2

Parameter estimates for the two-dimensional model of White participants' perceptions of female and male White, Black, and Asian faces.

	White Male		Black Male		Asian Male		White Female		Black Female		Asian Female	
Factor Loadings	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Trustworthiness dim.												
Caring	1.000	---	.990	.057	.898	.057	.990	.046	.851	.031	.907	..045
Warmth	.986	.050	1.000	---	1.000	---	.990	.045	1.000	---	.997	.044
Friendly	.992	.048	.895	.058	.922	.057	1.000	---	.920	.035	1.000	---
Attractive	.458	.091	.466	.069	.282	.071	.705	.093	.628	.070	.467	.083
Trustworthy	.471	.076	.600	.051	.674	.061	.784	.060	.619	.047	.605	.054
Aggressive	-.943	.065	-.867	.066	-.784	.064	-.821	.063	-.851	.044	-.769	.054
Healthy	.612	.098	.486	.081	.430	.075	.760	.085	.682	.068	.584	.077
Youthful	.522	.144	.320	.120	**.149**	.134	.558	.106	.452	.077	.450	.095
Dominance dim.												
Dominant	1.000	---	1.000	---	-.520	.106	1.000	---	**-.349**	.068	1.000	---
Physically Strong	.827	.057	.592	.073	-.409	.118	.735	.085	**-.056**	.060	.548	.091
Intelligent	**.026**	.065	**-.188**	.059	1.000	---	**.234**	.111	1.000	---	**-.161**	.091
Smart	**.034**	.065	**-.086**	.057	.933	.085	**.265**	.111	.873	.041	**-.204**	.086
Competent	**.222**	.063	**-.095**	.056	.821	.071	.500	.137	.903	.058	**.006**	.094
Assertive	.704	.042	.573	.064	**-.205**	.093	.982	.104	**-.049**	.063	.700	.095
Covariance:	*B*	*p*			*B*	*p*	*B*	*B*	*p*	*p*	*B*	*p*
Trustworthiness ~~ Dominance	**-.189**	<.001	**-.220**	<.001	**.194**	<.001	-.061	.059	**.463**	<.001	**-.199**	<.001

Note. In each model, the highest-loading trait on a dimension has an estimate (*B*) of 1.000 and no stand error (*SE*). General guidelines for interpreting the above estimates: Indicators that load strongly on its latent dimension have an estimate of > .800. Indicators that load moderately well have an estimate between .400-.800. Indicators that load weakly have an estimate < .400.
B = estimate. *SE* = stand error. *p* = p-value.
Traits which load poorly on its latent factor are **bolded**.
Significant (p < .05) covariances between the two latent dimensions are **bolded**.

While the parameter estimates do suggest some degree of covariance between the Trustworthiness and Dominance dimensions, these estimates are potentially unreliable. Furthermore, most of the traits that load poorly were specified to load on the Dominance dimension (Table 2). For instance, across most of the models, there did not appear to be much shared variance between intelligent/smart/competent and dominant/physically-strong. These weak loadings suggest that the inclusion of both sets of traits on the same latent dimension may be contributing to poor model fit. However, due to the unreliability of these parameter estimates, it is difficult to compare estimates and infer patterns across groups.

Overall, the present study was unable to reproduce the classic two-dimensional structure for both own-group and other-group perceptions, suggesting that the model may not be as universal as presumed. I discuss the implications of these findings in the Discussion section.

Three-Dimensional Models

As noted in the Introduction, recent work in the domain of face perception has extracted a third dimension of Youthful/Attractiveness (Sutherland, 2015; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016; Sutherland, Young, et al., 2016). Accordingly, I also modeled the three-dimensional structure of own-group and other-group faces. Again, six different models were built for White perceivers' ratings of female and male White, Black, and Asian faces. The specific trait loadings on each dimension were equivalent across all models

(Sutherland, 2015; Sutherland et al., 2013). See Figure 4 for a path diagram of this three-factor model.

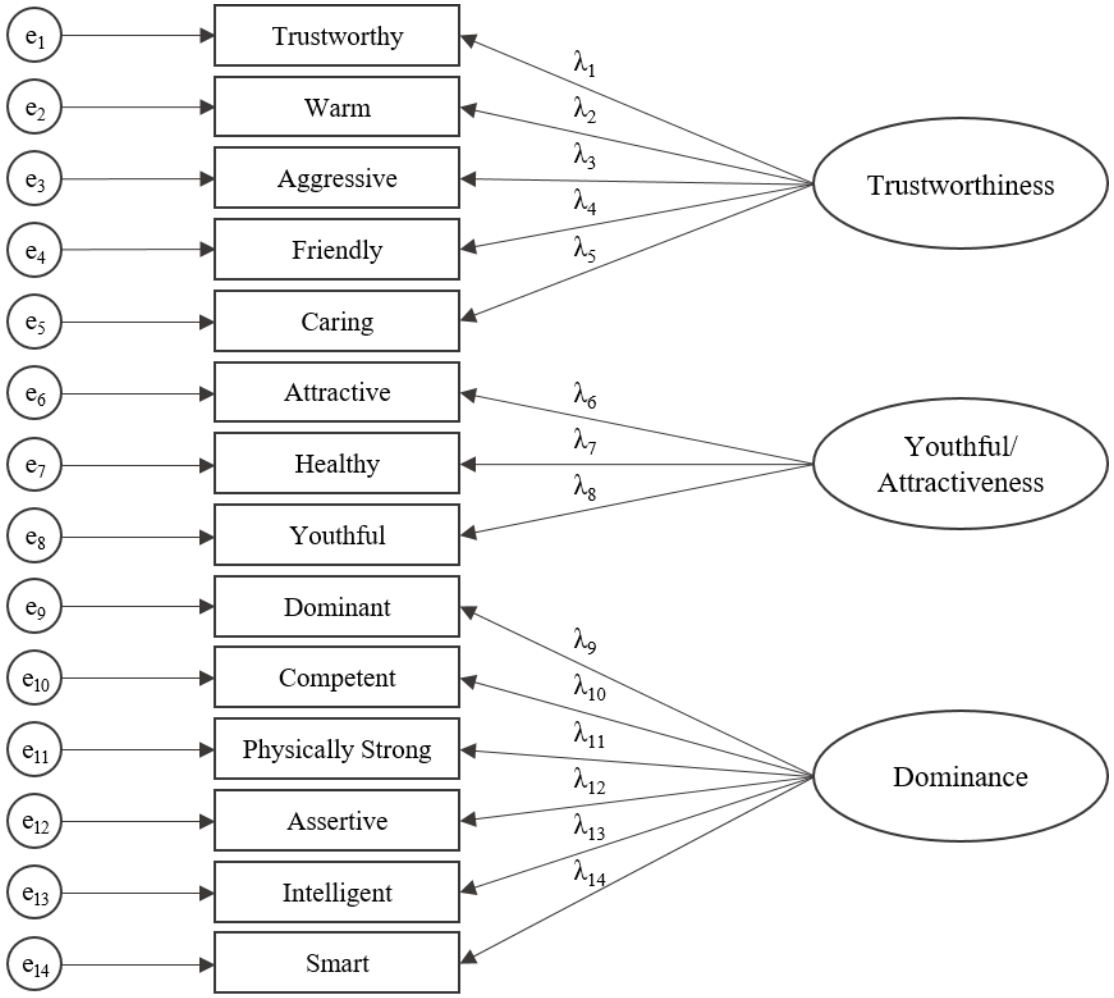


Figure 4. SEM path diagram for a three-dimensional CFA, representing the loadings of 14 traits onto their latent dimensions of Trustworthiness, Dominance, and Youthful/Attractiveness, which were fixed to be orthogonal. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait (e.g., trustworthy) served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable.

Model fit. We compared model fit indices across various own-group and cross-group perceptions, again expecting the three-factor structure to produce better model fit for own-group perceptions compared to cross-group perceptions. Similar to the previous analysis, however, results revealed poor model fit across all 6 models. While the three-dimensional models appear to be slightly better-fitting than the two-dimensional ones, with indices closer to the cut-offs, they still fall short of acceptance. See Table 3 for a full report of model fit indices.

Table 3

Indices of model fit for the three-dimensional model of face perception across own-group and cross-group perceptions.

	Chi-square ($\chi 2$)	*p* ($\chi 2$)	CFI	RMSEA	SRMR
Model					
White perceivers rating White males	987.56	<.001	.558	.287	.319
White perceivers rating Black males	801.96	<.001	.586	.256	.284
White perceivers rating Asian males	552.51	<.001	.682	.208	.158
White perceivers rating White females	552.98	<.001	.772	.208	.185
White perceivers rating Black females	508.20	<.001	.807	.201	.150
White perceivers rating Asian females	460.86	<.001	.754	.201	.182

Note. CFI = Comparative fit index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean squared residual.

Below are general rules for acceptable fit if data are continuous (Hu & Bentler, 1999; Schreiber et al., 2006):

1. Chi-square: Ratio of $\chi 2$ to *df* ≤ *2 or 3*.
2. CFI: ≥ .95 for acceptance.
3. RMSEA: < .06 to .08.
4. SRMR: ≤ .08

Although these indices are well below the acceptable limits, again, it is not possible to conduct traditional model modification to improve model fit, as these traits are spontaneously inferred from faces and cannot simply be removed from the model. Although I report the parameter estimates from the three-dimensional models in Table 4 below, in line with common reporting practices, these estimates should be considered less reliable due to poor model fit.

Table 4

Parameter estimates for the three-dimensional model of White participants' perceptions of female and male White, Black, and Asian faces.

	White Male		Black Male		Asian Male		White Female		Black Female		Asian Female	
Factor Loadings	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Trustworthiness												
Caring	1.000	---	.995	.057	.901	.056	.996	.046	.851	.030	.909	.043
Warmth	.981	.049	1.000	---	1.000	---	.992	.045	1.000	---	1.000	---
Friendly	.986	.048	.899	.058	.917	.057	1.000	---	.920	.033	.998	.044
Aggressive	-.951	.063	-.878	.066	-.790	.063	-.826	.063	-.842	.044	-.763	.054
Trustworthy	.466	.076	.599	.052	.664	.060	.777	.061	.610	.047	.598	.054
Youthful/Attractiveness												
Attractive	.712	.084	1.000	---	.571	.084	1.000	---	1.000	---	.867	.070
Healthy	1.000	---	.986	.106	1.000	---	.961	.057	.976	.060	1.000	---
Youthful	**.366**	.128	.673	.142	.866	.149	.657	.086	.664	.077	.626	.097
Dominance												
Dominant	1.000	---	1.000	---	-.465	.105	**.065**	.070	**0.342**	.067	**-.172**	.119
Physically Strong	.846	.061	.620	.073	**-.349**	.116	**.149**	.059	**-.052**	.060	**-.008**	.106
Intelligent	**.083**	.068	**-.205**	.062	1.000	---	.797	.044	1.000	---	1.000	---
Smart	**.089**	.068	**-.117**	.060	.941	.080	.827	.042	.862	.037	.918	.077
Competent	**.274**	.065	**-.119**	.059	.806	.068	1.000	---	.876	.057	.985	.085
Assertive	.758	.044	.595	.063	**-.158**	.091	**.191**	.070	**-.042**	.062	**.200**	.106
Covariance:	*B*	*p*			*B*	*p*	*B*	*B*	*p*	*p*	*B*	*p*
Trustworthiness ~~ Dominance	**-.179**	<.001	**-.225**	<.001	**.189**	<.001	**.340**	<.001	**.455**	<.001	**.208**	<.001
Trustworthiness ~~ Youthful/Attractive	**.233**	<.001	**.183**	<.001	**.130**	<.001	**.330**	<.001	**.423**	<.001	**.284**	<.001
Dominance ~~ Youthful/Attractive	**.159**	<.01	-.033	.276	**.158**	<.001	**.540**	<.001	**.487**	<.001	**.306**	<.001

Note. In each model, the highest-loading trait on a dimension has an estimate (*B*) of 1.000 and no stand error (*SE*). General guidelines for interpreting the above estimates: Indicators that load strongly on its latent dimension have an estimate of > .800. Indicators that load moderately well have an estimate between .400-.800. Indicators that load weakly have an estimate < .400.

B = estimate. *SE* = stand error. *p* = p-value.

Traits which load poorly on its latent factor are **bolded**.

Significant (p < .05) covariances between the two latent dimensions are **bolded**.

Similar to the two-dimensional models, there is significant covariance between the three dimensions. Notably, the direction of this relationship was positive for 4 of the 6 groups, but negative for Whites' perceptions of both White and Black men. Further, the magnitude of this covariance differs across groups. Also similar to the two-dimensional models, certain sets of traits (intelligence/smart or dominant/physically strong) load weakly on the Dominance dimension depending on whether 'dominance' or 'intelligence/competence' is the strongest indicator. These findings suggest that there may not be much shared variance between facial perceptions of competence/intelligence and dominance/physical strength. Again, however, I caution against over-interpreting these parameter estimates due to poor model fit.

One of the key limitations of CFA in an SEM framework is the difficulty of interpreting parameter estimates when we have poor model fit. In order to examine these trait relationships statistically and compare patterns across gender and racial groups, I turned to a different analytic approach.

Representational Similarity Analysis

The initial aim of this work was to examine whether the way in which humans evaluate faces varies across group boundaries. Due to the poor model fit of all models, we were unable to examine our hypotheses. To allow for some hypothesis testing, I turned to representational similarity analysis (RSA). The RSA allows for the comparison of trait matrices across different groups on the basis of some specific factor (e.g., race, gender), and tests my research questions in a different framework. Because the two- and three-dimensional models of face perception, with their poor fit indices, cannot be used to meaningfully describe the "trait space" that people use to evaluate own- and other-group targets, RSA provides a method of examining the trait matrices themselves across different groups. Specifically, are there associations between certain traits

(e.g., cross-loading between traits that load on different dimensions, or dissimilarity between traits that *should* theoretically load on the same dimension) that might explain the poor model fit? For instance, if the two sets of traits that load weakly on the Dominance dimension (intelligent/smart/competent and dominant/physically strong) do not share much variance in the trait space, an RSA would capture that dissimilarity. Furthermore, are certain traits more similar (or closely associated) for perceptions of specific racial or gender groups? If the similarity of trait pairs changes as a function of the target's race and gender, then it may indicate that perceptions of some groups may be more unidimensional than others.

To conduct RSA, the data were restructured such that each column was a correlation matrix of traits for a specific group. I created a 14 x 14 trait correlation matrix for each of the 6 groups (i.e., ratings of female and male White, Black, and Asian faces), removed all repeated trait pairs, and converted each matrix to "long" format (i.e., each column contained all trait-pair correlations for a particular group), producing 6 total columns of 91 trait-pair correlations each. Visualizations of trait correlation matrices for each group are provided in Figure 5. For descriptive purposes, I note that certain trait-pairs (e.g., competent and dominant) are weakly positively correlated in perceptions of some groups (e.g., White males), negatively correlated in others (e.g., Black males and females), and not at all correlated in others (e.g., Asian females). The RSA applies this comparison statistically to *all* trait-pairs across all groups, providing a test of whether trait-pairs are more dissimilar (or similar) as a function of target race (DV1), target gender (DV2), and a target race × gender interaction.

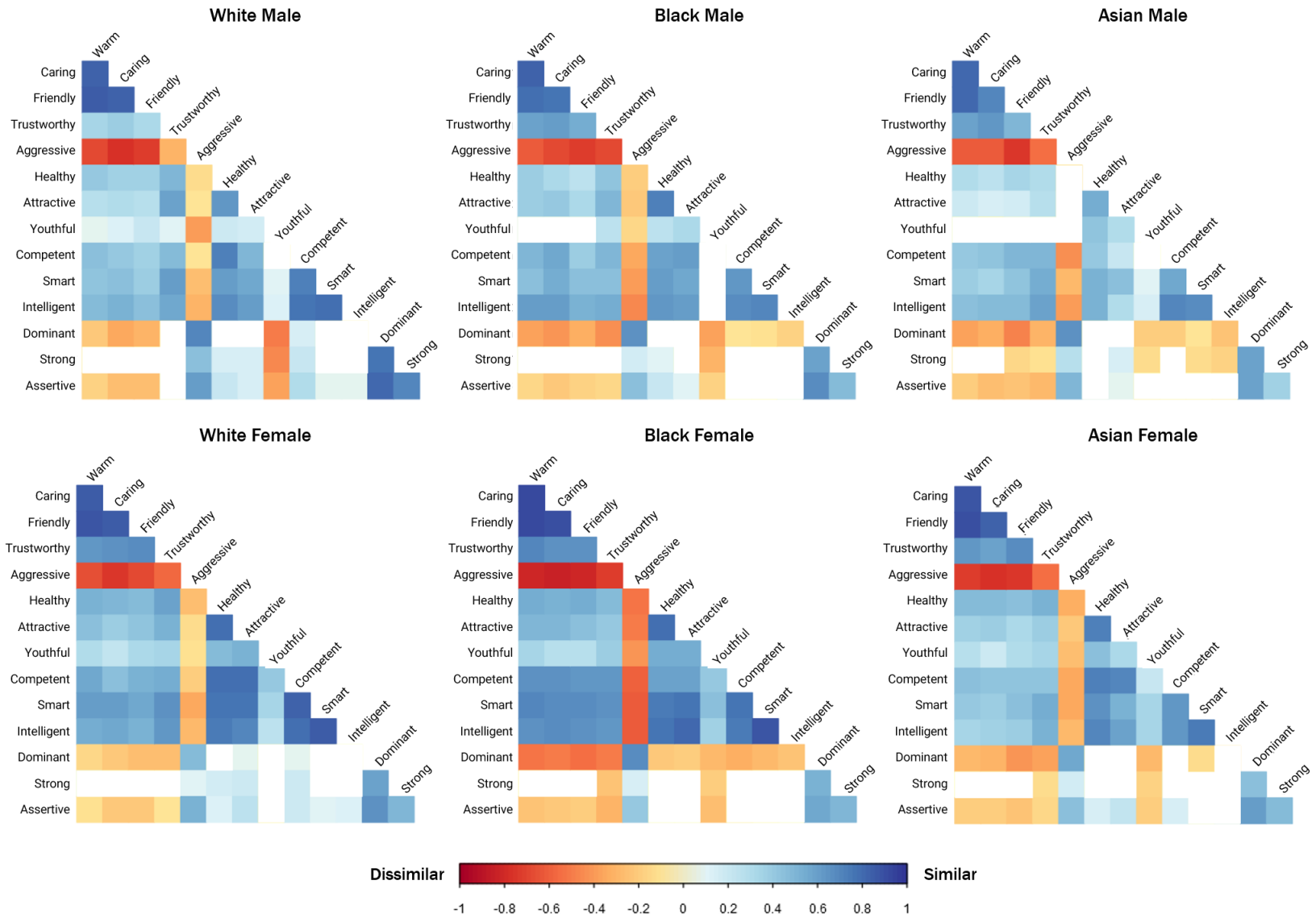


Figure 5. Correlation matrices of 14 traits for White participants' perceptions of female and male White, Black, and Asian faces. Darker colours represent stronger correlations (closer to $r = 1$ or $r = -1$) between pairs of traits. Blank squares indicate trait-pairs that are non-significant at $p < .05$.

To perform RSA in the present study, I conducted an RMANOVA on the 6 different trait correlation matrices, in a 2 (Target Gender: Female, Male) × 3 (Target Race: White, Black, Asian) design. Analyses were conducted in R using the *anova* (Fox & Weisberg, 2011) and *ezANOVA* (Lawrence, 2016) packages, as well as in SPSS.

Descriptive statistics. Across 6 groups, the 14 traits (i.e., 91 trait-pair correlations) were weakly but positively correlated. Perceptions of White female faces had the highest mean correlation across all trait-pairs (M_r = .300, *SD* = .422), followed by White males (M_r = .232, *SD* = .418), Asian females (M_r = .217, *SD* = .431), and Black females (M_r = .208, *SD* = .530). Black males (M_r = .198, *SD* = .431) and Asian males (M_r = .156, *SD* = .404) had the lowest average correlation across trait-pairs.

Mauchly's test, $\chi^2(2) = 14.56$, $p = .001$ indicated minor violations of sphericity. As such, I report the corrected output from the within-subject tests below.

Within-subjects effects. Results indicate that correlations between trait-pairs were *not* equal across race and gender boundaries. There was a significant main effect of Target Gender on the correlations of trait-pairs, $F(1,91) = 15.81$, $p < .001$ and a significant main effect of Target Race, $F(2,91) = 26.25$, $p < .001$, both qualified by a significant Gender × Race interaction, $F(2,91) = 5.88$, $p < .01$. There were no differences in *p*-values across sphericity-corrected (i.e., Greenhouse-Geisser, Huynh-Feldt, Lower-bound) vs. non-corrected outputs, indicating that violations of normality did not change conclusions. A table of the estimated marginal means (by Target Gender, Race, and the Gender × Race interaction) are provided in Table 5.

Table 5

Estimated Marginal Means from an RMANOVA on the 6 trait-pair correlation matrices in a 2 (Target Gender) × 3 (Target Race) design.

	Mean	*SE*	95% CI LL	95% CI UL
Measure				
Gender				
Female	.242	.048	.147	.337
Male	.195	.043	.110	.280
Race				
Asian	.186	.043	.101	.272
Black	.203	.050	.104	.302
White	.266	.043	.180	.352

Gender × Race				
Asian Female	.217	.045	.127	.307
Black Female	.208	.056	.098	.319
White Female	.300	.044	.212	.388
Asian Male	.156	.042	.071	.240
Black Male	.198	.045	.108	.287
White Male	.232	.044	.145	.319

Overall, these results indicate that, on average, the *association* between different pairs of trait ratings (e.g., competent, attractive) differs across groups as a function of race and gender. Critically, because the unit of analysis is the correlation of trait-pairs, the direction of these associations can be positive or negative. For instance, if ratings of "warmth" and "competence" are highly dissimilar (negatively correlated) for perceptions of White females but highly similar for White males, then this difference will have been captured by the RSA. These findings provide support for the hypothesis that evaluations of some groups are more unidimensional than others, given that the similarity (or dissimilarity) of different trait pairs changes significantly depending on the race and gender of the target group. Finally, although I directly tested the unidimensional model (i.e., a single-factor model with all 14 traits loading on one factor) across all 6 groups using CFA, these models again had extremely poor fit. The results of this one-factor model are reported in the Appendix.

While the RSA technique conducted in an RMANOVA framework does not reveal which specific groups have more similar trait spaces (i.e., are evaluated more unidimensionally) than others, nor the direction of correlations between trait-pairs, it provides a useful supplement to the findings from CFA.

General Discussion

Social impressions are powerfully influenced by faces. Within milliseconds of perceiving a face, we spontaneously infer enduring personality traits about a person, such as how friendly or competent they are (Oosterhof & Todorov, 2008; Todorov et al., 2015; Willis & Todorov, 2006). Researchers in the domain of face perception have used data-reduction techniques to distill these myriad social impressions down to two latent, independent dimensions: Trustworthiness and Dominance (Oosterhof & Todorov, 2008; Todorov, Baron, et al., 2008; Todorov et al., 2013; Todorov, Said, et al., 2008), with recent work finding a third dimension of Youthful-Attractiveness (Sutherland, 2015; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016). While these prominent models of face and person perception are presumed to be universal, reflecting how humans perceive and evaluate all other humans, the generalizability of these models have not been rigorously scrutinized. Within the face perception domain, the two-dimensional model was developed and validated with White participants' ratings of computer-generated White male faces (Oosterhof & Todorov, 2008; Todorov et al., 2013) – and whether this model generalizes to cross-group perceptions of real facial stimuli has yet to be examined.

The influence of group identity on social perceptions is at the theoretical heart of social cognition and intergroup research. Across a substantial body of work, evidence suggests that the two dimensions of Trustworthiness and Dominance are *not* independent for cross-group perceptions (Cuddy et al., 2009; Eagly & Kite, 1987; Glick & Fiske, 1996; Holbrook et al., 2016; Livingston & Pearce, 2009; Wilson et al., 2017). Given extant evidence that trait impressions across racial or gender groups may be more strongly correlated, or occur in a more unidimensional manner in which judgments on one dimension bias the other, testing these models with both *own-group* and *other-group* perceptions was the focus of the current work. I argue that classic models of face perception do not generalize across race or gender groups: (1)

results indicate that these models fit poorly across all groups, and (2) although the parameter estimates for these models are potentially unreliable, and should not be over-interpreted, the dimensions of Trustworthiness and Dominance do vary in the extent to which they correlate as a function of targets' race and gender, and are not always independent.

In the current work, I took a rigorous multi-measurement approach to testing current models of face perception. Conducting CFA in an SEM framework, I fit the classic two- and three-dimensional models to ratings of faces (varying along race and gender) across 14 traits demonstrated to emerge spontaneously in facial first impressions and to load on their theoretical dimensions. Specifically, I built separate factor structure models for *own-group* (i.e., White participants rating White female or male targets) and *other-group* (i.e., White participants rating Black or Asian female or male targets) perceptions, and examined model fit indices for each model to assess goodness-of-fit. Thus, for perceptions across group boundaries, I assessed how well the theoretical covariance matrix (defined by the orthogonal two- and three-factor models) mapped onto the observed covariance matrix. Further, I examined parameter estimates of these models to determine whether the degree of correlation between Trustworthiness and Dominance varied as a function of the target's race and gender. In a subsequent analysis, I conducted a representational similarity analysis (RSA) of the trait space represented by each model (i.e., a correlation matrix of traits for each race/gender group), to determine whether the *similarity* of trait spaces (i.e., the strength of association between a pair of traits across all trait-pairs) differed across groups as a function of target race and gender. Here, the aim was to test whether the space of trait-pairs are more similar or dissimilar for perceptions across group boundaries. If the similarity of trait pairs changes across racial and gender groups, then it would indicate that

perceptions of some groups may be more unidimensional (i.e., more correlated overall) than others.

Critically, the present work demonstrates that current dominant models of face perception produce poorly-fitting models across *all* groups, which indicates that the classic two- and three-dimensional structure of face perception did not generalize to this dataset of own-group and cross-group perceptions. By conducting CFA in an SEM framework, I was unable to replicate the exploratory two-dimensional structure of Trustworthiness and Dominance developed in face perception research. Further, results from the similarity analysis indicated that the conceptual space of trait associations changes significantly across race and gender categories. For instance, ratings on a pair of traits (e.g., competent and dominant) may be similar for one group (e.g., White males) but unrelated for another (e.g., Asian females). Thus, the similarity of trait representations differs as a function of target race and gender, indicating that trait perceptions of some groups may be more strongly correlated than others.

Overall, the findings highlight the importance of validating data-driven models with multiple measurement frameworks, as well as the limitations of current models of person perception in generalizing across racial and gender groups. Future research in developing a universal model of person perception that generalizes across different stimuli and groups should attempt to explore group differences in the trait space of face judgments. That is, if differences in the representational similarity of a pair of traits (e.g., competent and dominant) across groups (e.g., White males and Asian females) arise from some third factor (e.g., stereotypes, perceived threat, evaluative context, perceiver characteristics, target morphological features, etc.), then this third factor should be modeled to develop an adequately-fitting universal model of face perception. In SEM frameworks, comparisons between the theoretical and observed covariance

matrices are restrictive. Thus, without including this "third" factor in the model, it may not be possible to achieve adequate fit with current two- and three-dimensional models. I elaborate on this perspective in the following sections.

Classic Models of Person Perception

The observed model tested in this thesis did not validate Todorov and colleagues' (2008a, 2008b, 2013) theoretical two-dimensional model of face perception. Across all groups, model fit indices fell short of their acceptance ranges, and parameter estimates indicated correlations between the two dimensions previously thought to be independent. Thus, our test of Oosterhof and Todorov's (2008) exploratory model of facial first impressions, which posits that people evaluate others on two independent dimensions of Trustworthiness and Dominance, was not supported for either own-group or other-group perceptions.

Given our large sample size, it is unlikely to be a factor in the failure of two-dimensional models to reach acceptable fit. Although there are no exact rules for sample size determination in SEM – the great variability in sample size requirements makes it difficult to apply a "one size fits all" approach (MacCallum et al., 1996) – the general consensus recommends 10 cases per estimated parameter (for a review, see Schreiber et al., 2006). However, recent simulation studies suggest that smaller sample sizes may be adequate, depending on the specification of one's model (Wolf et al., 2013), the number of latent factors (more factors require larger *N*), and the number of indicators per latent factor (fewer indicators require larger *N*). For instance, Wolf and colleagues (2013) recommend ~150 cases per indicator for a two-factor model with >6 indicators loading on each factor, when there are no missing cases in the data and when indicators are expected to load moderately well. Here, each case is the average rating of a stimulus on a particular trait across all participants who rated that stimulus on that trait, and there are 14

parameters (i.e., trait indicators). Since there are ~150 unique stimuli for each group (with the exception of Asian females, for which there are only 129), I was usually able to achieve an acceptable ratio of 10.7 cases to 1 parameter estimated. Further, because the unit of analysis is the averaged stimulus rating, there are no missing cases in the data structure. Finally, it is important to note that at the time of data collection, the sample size was determined based on requirements for the two-factor model, and may be insufficient for estimations of the three-factor model. This limitation is given greater consideration in subsequent sections that focus on the three-factor model.

Given adequate sample size, which critically relates to the stability of parameter estimates, I turn to other reasons for the poor fit of the hypothesized two-dimensional model. One limitation of CFA is its reliance on the highly restrictive independent cluster framework, in which cross loadings between indicators and non-target factors are fixed to zero. This is seldom the case with real data, as many traits do cross-load in reality. For instance, the 'attractive' trait correlates positively with a host of other traits from other dimensions, such as competence, intelligence, and health, due in part to the halo effect (Eagly, Ashmore, Makhijani, & Longo, 1991; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Zebrowitz, Hall, Murphy, & Rhodes, 2002; Zebrowitz & Rhodes, 2004). In a CFA framework, however, indicators that are not specified to load on the same target factors (e.g., attractive and competent load on Trustworthiness and Dominance respectively) are not allowed to share variance, typically resulting in poor model fit.

One solution might be to test the models in a bifactor exploratory SEM framework, in which indicators that are not a priori set to load on the same target factors are allowed to share variance. The bifactor model would allow researchers to examine the *unique* variance of each

indicator given all other indicators – in other words, all of the shared variance between indicators that are not being captured by the two dimensions are extracted into a third general, unspecified latent factor. This general factor may represent the halo effect, or some other dimension, or possibly even a non-interpretable "junk" factor that uniquely explains the leftover shared variance across all indicators (Morin et al., 2016). While these bifactor SEM models may improve model fit by allowing indicators to cross-load, they also require extremely large sample sizes to successfully converge. Furthermore, without proper theoretical grounding, the third general factor extracted from these exploratory models may be difficult to interpret, as it may simply represent a junk factor that captures all of the leftover variance shared across indicators. More consideration is given to similar exploratory approaches in later sections.

There may also be theoretical reasons for the overall lack of model fit. Examining the parameter estimates for each race/gender model reveals broader patterns that can inform our understanding of the underlying trait space, and how it maps onto the theoretical dimensions. First, although several indicators on the Trustworthiness dimension have moderate to strong loadings (i.e., warm, friendly, caring, and trustworthy), other indicators such as attractive, healthy, and youthful load weakly onto this dimension across all groups. Consistent with recent work by Sutherland and colleagues (2013, 2015), which identified three dimensions of face perception, these traits may be better captured by a third dimension of Youthful-Attractiveness. These studies had incorporated more diverse and naturalistic facial stimuli with a broader age range compared to Oosterhof and Todorov's (2008) exploratory work on the two-dimensional model, which may explain why a third dimension had been extracted. In the present work, our findings tentatively provide some support for the three-factor vs. two-factor model of face perception. However, I note that while the three-dimensional models had slightly better model fit

indices than the two-dimensional models, they still fell short of acceptance. As such, we were unable to confirm the generalizability of the three-dimensional model.

More strikingly, several indicators on the Dominance dimension load poorly in systematic ways. In each model, the highest-loading indicator on a factor is assigned a loading of 1.000, such that the parameters of all other indicators on that factor are estimated from this primary indicator. Across race/gender groups, the specific trait selected to be the primary indicator vacillates between dominant and intelligent, and produces strong loadings on 2-3 related indicators but weak loadings on the others. For instance, intelligent and smart (and, for certain groups, competent) are linked such that they have similar loading strength on the factor – smart loads strongly if intelligent is the primary indicator, and weakly if dominance is the primary indicator. The other set of indicators, dominant and physically strong, are similarly linked. This divergence between intelligence/smart and dominant/strong suggests that ratings on these traits do not consistently share much variance. Indeed, an examination of the correlation matrix in Figure 5 reveals weak or non-significant correlations between these pairs of traits for some groups, and negative correlations for others.

Competence vs. dominance. If dominant/strong and intelligent/smart/competent do not share much variance, then a model in which they are specified to load on the same dimension will invariably fit poorly. Should these traits load on the same dimension? Trustworthiness (cf. Warmth) and Dominance (cf. Competence) are two dimensions that emerge consistently across the face perception and intergroup perception literatures (Fiske et al., 2007, 2002; Oosterhof & Todorov, 2008; Sutherland et al., 2013), but despite their apparent parallels and shared theoretical underpinnings, these dimensions may not be equivalent. Recent work integrating social and facial models of person perception found key areas of divergence (Sutherland,

Oldmeadow, et al., 2016). Specifically, whereas perceptions along the first dimension of Trustworthiness or Warmth are closely related, demonstrating strong similarity across models varying in stimuli (e.g., ratings of faces vs. abstract judgments of people or groups), perceptions along the other dimension of Dominance or Competence are less similar. Critically, the divergence between Dominance and Competence may reflect different routes to evaluating a target's "ability" to enact their intentions (Sutherland, Oldmeadow, et al., 2016).

Indicators of a target's ability may vary across contexts, such as the stimulus being evaluated (e.g., a face or a social group), the evaluative context (e.g., success in politics or the military; Ballew & Todorov, 2007; Chen et al., 2014; Hehman, Leitner, et al., 2015; Mueller & Mazur, 1996), or even the social group to which the target belongs (e.g., a female or male face; Sutherland, Oldmeadow, et al., 2016). For instance, the relationship between competence and dominance was weaker for ratings of female faces than for male faces (Sutherland, Oldmeadow, et al., 2016). While our findings provide some support for this line of work, in that the specific traits which load well onto the "dominance" dimension changes across racial and gender groups (e.g., "dominant" is the primary indicator loading on its latent factor for perceptions of White men, but "intelligent" is the primary indicator for Asian men), these parameter estimates are difficult to interpret because model fit is so poor. Instead, I turned to RSA to statistically compare the underlying trait space across different racial and gender groups.

Representational Similarity of the Social Trait Space Differs Across Groups

The initial aim of this thesis was to examine whether correlations between the dimensions of face perception (i.e., Trustworthiness, Dominance) varied across group boundaries. Due to the poor model fit of all two- and three-factor models, I was unable to examine this hypothesis using CFA. To allow for some hypothesis testing, I turned to representation similarity analysis (RSA)

to examine whether the similarity of trait-pairs in the underlying trait space of facial perceptions changes across groups. Specifically, does the correlation between different pairs of traits, across all trait-pairs, differ as a function of race and gender? If so, it would suggest that some groups are perceived more unidimensionally than others, given that trait ratings of these group are more similar overall.

Results reveal that, on average, the similarity between different pairs of trait ratings (e.g., competent, attractive) varies across groups as a function of race and gender. We found a main effect of Race and Gender, as well as a significant Gender × Race interaction, on the correlations between each trait-pair. Thus, the associations between trait-pairs in the underlying "conceptual trait space" are more similar for some groups than others, providing support for the hypothesis that evaluations of some groups are more unidimensional than others. It is important to note, however, that this analysis does not reveal which specific groups have more similar trait spaces than others – just that the similarity of trait spaces varies significantly across groups. In the next paragraphs, I offer some speculation on the theoretical basis of these group differences, though more research is needed to adjudicate between various theories.

Group differences in the social trait space can stem from a variety of factors, such as stereotypes, evaluative context, and group processes. In the face perception domain, recent research has found gender stereotypes to influence judgments on specific traits in systematic ways. For example, perceptions of trustworthiness and dominance are more negatively related in female vs. male faces (Sutherland et al., 2015), in line with group stereotypes about traditional women (i.e., homemakers) being perceived as warmer but less competent, whereas professional women are colder but more competent (Cuddy et al., 2004; Fiske et al., 2002, 1999; Ruble, 1983; Sutherland, 2015). When professional women become mothers, they trade perceived competence

for perceived warmth, whereas professional men who become fathers maintain perceived competence while gaining perceived warmth (Cuddy et al., 2004). In a similar vein, the stereotype content model (SCM) posits that White males are a trusted in-group in the US, and are typically rated as both warm and competent.

Group stereotypes can also interact with social motivations to shift evaluations on the trait space. Perceivers may construe dominance/competence traits as positive for close and trusted groups (e.g., White males) but as negative and threatening for outgroups (Fiske et al., 2002; Hutchings & Haddock, 2008; Kubota & Ito, 2017; Sinclair & Kunda, 1999; Stolier et al., 2018). Accordingly, trustworthiness and dominance are positively related in perceptions of close and admired targets but negatively correlated in perceptions of distant or outgroup targets (Fiske et al., 2007, 2002; Lin, Kwan, Cheung, & Fiske, 2005). For example, Black compared to White men tend to be stereotyped as bigger and more physically threatening (Wilson et al., 2017), and accordingly, counter-stereotypical Black men who look more baby-faced (i.e., less physically dominant) are perceived as warmer and receive more rewards (Livingston & Pearce, 2009). In a similar vein, when participants are given information about a Black or White target's apparent intentions (Sinclair & Kunda, 1999, 2000), those who received positive feedback from a Black professional (i.e., indicating positive intentions) rated the target as more competent, and those who received negative feedback (i.e., indicating negative intentions) rated the target as less competent. In this case, perceptions of positive intentions buffer against the negative stereotypes associated with an outgroup member.

While descriptive, our findings are generally consistent with the idea that group stereotypes shift evaluations on the trait space. In our visualization of trait matrices across different race/gender groups (Figure 5), we note several patterns that are consistent with this

body of research. For instance, White participants' ratings of trustworthiness and dominance are unrelated for White male targets (i.e., the in-group), but are negatively correlated for all other groups (i.e., societal out-groups). Furthermore, although White females are considered an in-group, gender stereotypes about women as warmer or more submissive but less dominant could have shifted the trait space, such that trustworthiness and dominance become negatively correlated for White females. In a similar vein, whereas perceived warmth and dominance are negatively correlated across all groups, this association is particularly strong for Black males and females, in line with racial stereotypes. However, I reiterate that these specific cross-group patterns are descriptive. Our findings from the RSA indicate that the social trait space shifts as a function of race and gender, but it does not provide information about which groups are shifting in which direction, nor which trait-pairs are significantly different. Future research with planned contrasts between specific groups is required to statistically compare these differences in trait spaces.

To summarize, a general finding of the current work is that the two- and three-dimensional model of face perception does not generalize across racial and gender boundaries. We were unable to confirm Todorov and Oosterhof's (2008) exploratory two-dimensional model of Trustworthiness and Dominance, nor Sutherland and colleagues' (2013, 2015) three-dimensional model. Furthermore, we found that across both own-group and other-group perceptions, these dimensions were significantly correlated, and not independent as presumed. Furthermore, we specifically tested the similarity of the social trait space across all groups, and found that the representational similarity of traits (e.g., whether perceptions of 'trustworthy' are similar/related to perceptions of 'dominant') differed significantly as a function of targets' race and gender. Finally, it is important to note that while the present research satisfies the question of

whether current models of face perception are universal (i.e., they are not), future research is needed to examine *how* the trait space shifts across race and gender groups. With the two statistical approaches laid out here, researchers can disentangle the differences in trait space that are unique to each group in their own lines of work, and advance understanding of how race and gender interact to drive social impressions.

Limitations and Future Directions

There are several limitations of the present work. First, although our analyses included a large number of ratings of trait impressions (402,473 ratings of 873 unique stimuli across 5,040 participants), analyses in SEM were conducted with averaged ratings of each stimulus as the unit of analysis. Effectively, our sample size was the number of unique stimuli we had in each model (i.e., ~150 stimuli for each model). While, for most models, this number satisfied the sample size rule-of-thumb of 10 cases (~150 stimuli for each racial/gender group) to 1 parameter (14 trait indicators), we had fewer Asian female stimuli (only 129 unique stimuli) – potentially affecting the reliability of the output for perceptions of Asian females. Recent work in the area of structural equation modeling has provided tutorials on building multilevel structural equation models, and could, in future work, be extended to these analyses (Rabe-Hesketh, Skrondal, & Zheng, 2008; Sterba et al., 2014). For example, participants' ratings, which are nested within each unique stimulus, could serve as the unit of analysis – which would greatly increase our sample size.

Another limitation of CFA is the difficulty of interpreting parameter estimates when models fit poorly. These parameter estimates are of great interest in the present work, but may be potentially unreliable. Furthermore, CFA is limited by its highly restrictive framework, which fixes cross loadings between indicators and non-target factors to zero. Many traits cross-load in

reality, but are fixed to be independent in these models – likely contributing to poor model fit. One solution would be to test these models in a bifactor exploratory SEM framework, in which the unique shared variance between indicators (that are not being captured by the two dimensions) are extracted into a general, unspecified latent factor. These SEM models may improve model fit by allowing indicators to cross-load, and future work in this area could focus on testing whether a third factor (representing the halo effect, or Youthful-Attractiveness) effectively captures the leftover variance shared across indicators.

While this research comprises a test of current two- and three-factor models of face perception, not all traits submitted to our model were used in Todorov and Oosterhof's (2008) original models. We included traits that were semantically similar to other traits in their model, were present in subsequent replications of Oosterhof and Todorov's model (Sutherland et al., 2013; Todorov et al., 2013, 2011), and were found to spontaneously arise in social impressions (Asch, 1946; Fiske et al., 2007; Rosenberg et al., 1968; Sutherland et al., 2013; Sutherland, Oldmeadow, et al., 2016; Todorov et al., 2015; Zebrowitz & Montepare, 2008). Furthermore, although the present work examined own-group and other-group perceptions, all analyses were conducted with ratings from White participants, as we did not have enough ratings from Black or Asian participants. Future work testing the changes in trait spaces across groups should also examine Black and Asian perceptions of own-group and other-group targets.

Another limitation of the study concerns the variability in stimuli sources across different models. Past studies have shown that the type of facial stimuli (e.g., diverse vs. homogenous, controlled vs. naturalistic) used in these models can considerably influence which trait impression spontaneously arise, how much variance exists around these trait impressions, and even the number of dimensions extracted (Sutherland et al., 2013; Todorov et al., 2015; Vernon,

Sutherland, Young, & Hartley, 2014). I followed recommendations from previous research on data-driven approaches with heterogeneous ambient stimuli, and incorporated large stimuli samples from controlled databases that were diverse in their presentation and in their representation of different traits (Burton et al., 2015; Hehman et al., 2017; Jenkins et al., 2011; Sutherland et al., 2013; Vernon et al., 2014). However, the proportion of stimuli sourced from each database was not equal across groups. For instance, certain databases such as the Chicago Face Database (Ma et al., 2015) had fewer stimuli for some groups (e.g., Asian female, Asian male) than others. In order to meet the sample size requirements of SEM, I sourced the remaining stimuli from other databases. Thus, while all of the stimuli were sourced from highly controlled, standardized databases, there may still be differences between stimuli sets that are contributing to differences in trait spaces across groups. To the extent that the stimuli involved in the present research are not representative for any group or analysis, our effects may differ from future research using other stimuli.

Finally, although our findings from RSA demonstrated that the similarity of trait spaces shifts across race and gender groups, we did not examine which specific groups had differing trait spaces, nor which trait-pairs were driving these differences. Portions of this research describing different patterns in trait spaces were descriptive in nature. These results lay the foundation for future researchers to use the RSA approach to examine group differences in trait spaces, and researchers may test these results with confirmatory approaches. For example, future research using planned contrasts in RMANOVA can examine *how* the trait space shifts across race and gender groups, and inform our understanding of how to build a better, more generalizable model of face perception. By extending the RSA technique to other types of data beyond face ratings, we can examine how the conceptual space for other constructs (e.g., group

stereotypes) maps onto the face-based trait space, and whether this mapping differs across groups. For instance, researchers can collect data on participants' stereotypes of various social groups (e.g., rating Black women on warmth, dominance, etc.), create a group-stereotype trait matrix, and compare that with the face-based trait matrix to determine whether shifts in the group-stereotype matrix (e.g., warmth and dominance are negatively correlated) map onto shifts in the facial-trait matrix. Researchers would then be able to draw conclusions about the source of these group-based differences in the trait space, by demonstrating, for example, that the negative relationship between warmth and dominance in ratings of Black female faces arises due to the negative relationship between warmth and dominance from racial stereotypes.

Conclusion

In summary, the present work advances the impression formation and intergroup perception literatures by testing prominent models of face perception in a rigorous, confirmatory framework, and examining whether these models generalize to perceptions across race and gender groups. These results demonstrate that current models of face perception fit poorly across all groups, and that the two-dimensional model of Trustworthiness and Dominance derived using principle components analysis did not generalize to own- or other-group perceptions in an SEM framework. Further, across both own-group and other-group perceptions, these dimensions were significantly correlated, and not independent as presumed.

Using an RSA approach, this thesis also tested the similarity of the social trait space across all groups, and found that the representational similarity of traits (e.g., whether perceptions of 'trustworthy' are similar/related to perceptions of 'dominant') differed significantly as a function of targets' race and gender. These findings demonstrate that race and gender interact to influence how similar (or correlated) trait impressions of other individuals are.

Thus, the present work advances scholarly understanding of how to better model trait impressions across group boundaries, and provides some insight into the very nature and process of how people form impressions. Here, I demonstrate that current models of face perception are not universal, and reveal systematic differences in the conceptual trait space across group boundaries. Future research is required to determine how the trait space shifts across race and gender groups. With the two statistical approaches laid out in this thesis, researchers can explore the sources of differences in trait spaces across groups, and advance our understanding of how race and gender interact to drive social impressions. Our findings reveal the profound variance that exists in impression formation across groups, and demonstrates how a rigorous statistical framework can be used to test data-driven models of face perception – as well as how an emerging statistical approach can be used to test and compare the representational similarity of trait spaces across groups.

Appendix A: Study Materials

A.1 Consent Form on Amazon Mechanical Turk

We are conducting an academic survey about how humans form impressions of others. Below is the informed consent. Select the link below to complete the survey. At the end of the survey, you will receive a code to paste into the box below to receive credit for taking our survey.

You are being asked to participate in a research study. Before signing this consent form, it is important that you read the following information. You may ask as many questions as necessary to be sure that you understand what the study entails.

Investigators: Seeing Human Lab, Department of Psychology, Ryerson University.

Purpose of the Study: The purpose of this study is to examine how humans form perceptions of others.

Description of the Study: If you decide to participate in the research, you will be asked to view and rate images of faces, bodies, and objects that you would encounter in everyday life.

Risks or Discomforts: This is a minimal risk study. Participants may withdraw from the study at any time by closing their browser.

Benefits of the Study: There is no direct benefit to participants in this study although the information gained from the overall study may improve our understanding of how humans perceive others. When the session is over, you will have the opportunity to learn more about the purpose and hypotheses of the study.

Confidentiality: You will not be identifiable from your responses in this task, or from data recorded during your session. Your IP address will be recorded, which can be used to identify your country, region, and sometimes city. The data from this study will be held on a password protected hard drive to which only senior research personnel have access. De-identified data will be kept indefinitely. This survey or questionnaire does not ask for personal identifiers or any information that may be used to identify you. However, if you choose to participate in this task, you understand that your responses to the survey questions will be stored, and can be accessed, in the USA.

Incentives to Participate: You will be compensated for your time with payment listed on Amazon Mechanical Turk.

Voluntary Nature of Participation: Participation in this study is voluntary. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty.

Questions about the Study: If you have any questions about the research now or after the completion of the study, please contact [seeinghumanlab@gmail.com].

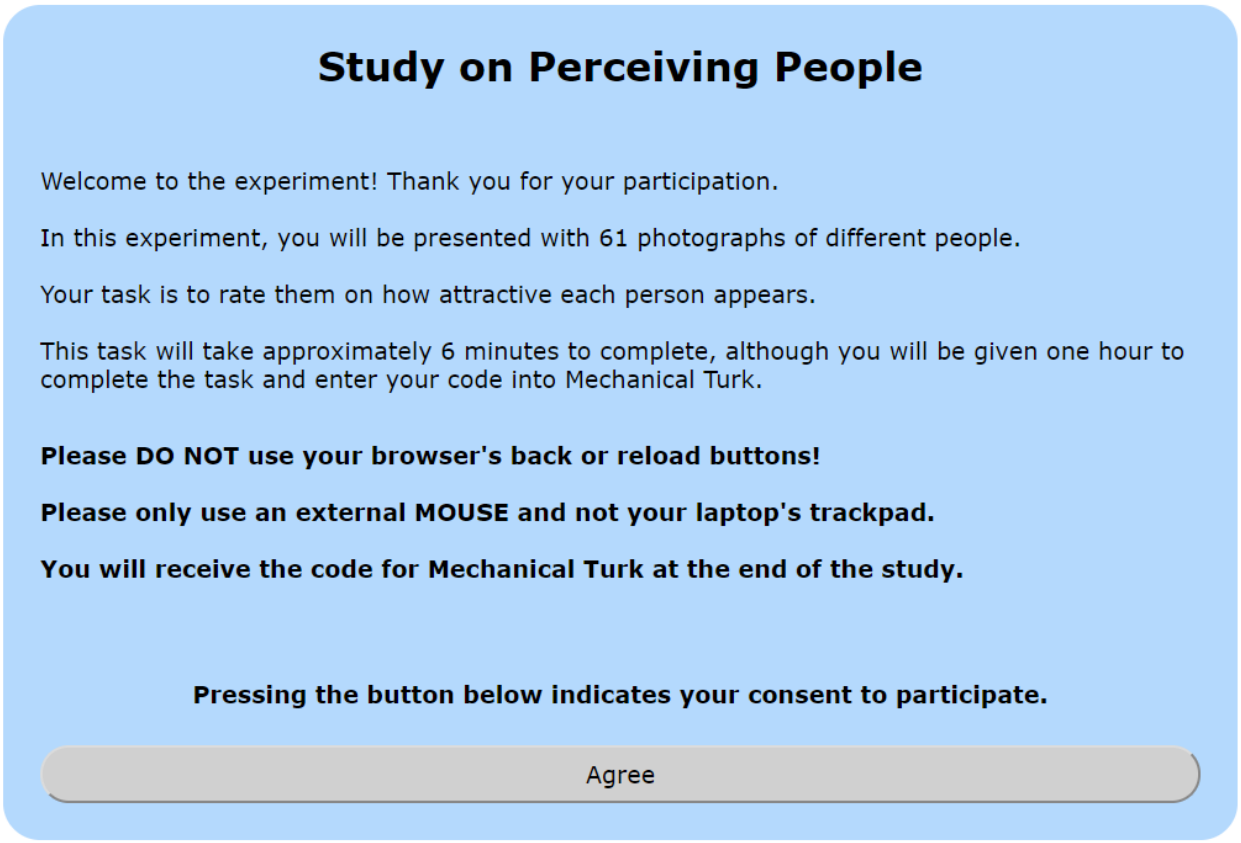
If you having any questions regarding your rights as a human subject and participant in this study, you may contact the Ryerson University Research Ethics Board for

information: Ryerson Ethics Board, c/o Office of the Vice President, Research and Innovation, Ryerson University 350 Victoria Street Toronto, ON M5B 2K3, 416-979-5042.

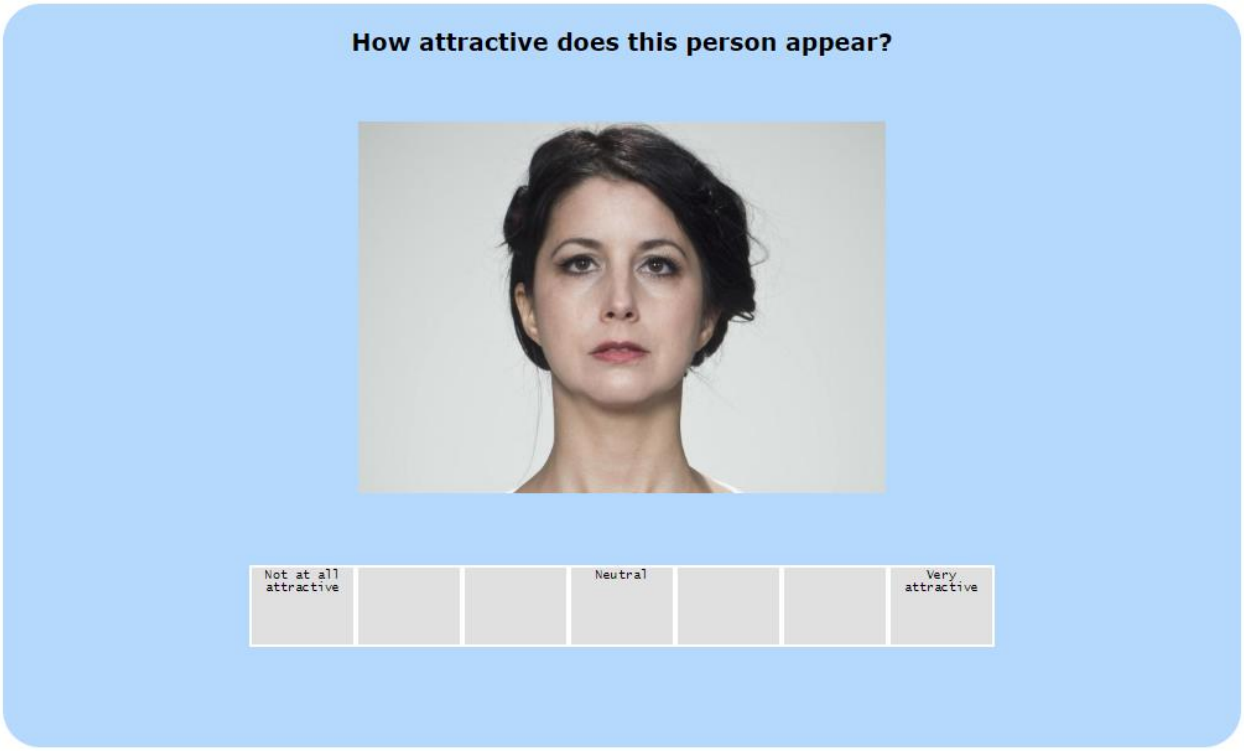
Agreement: By clicking the link to the task below you are indicating that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Continuing also indicates that you agree to be in the study and have been told that you can change your mind any time during the study and withdraw from it. Finally, continuing indicates you have been told that by signing this consent agreement, you are not giving up any of your legal rights.

Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box.

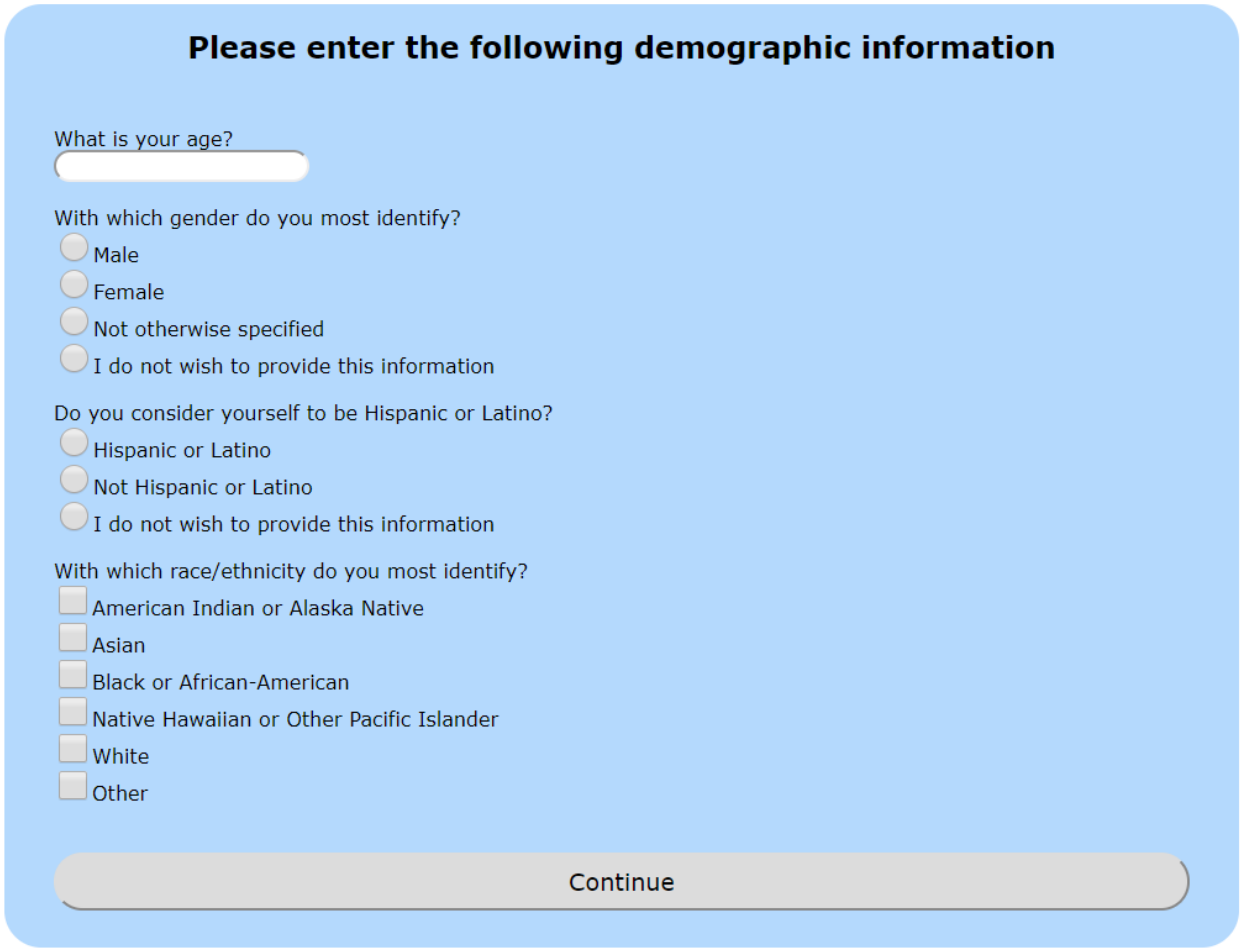
A.2 Sample Study Portal on Website



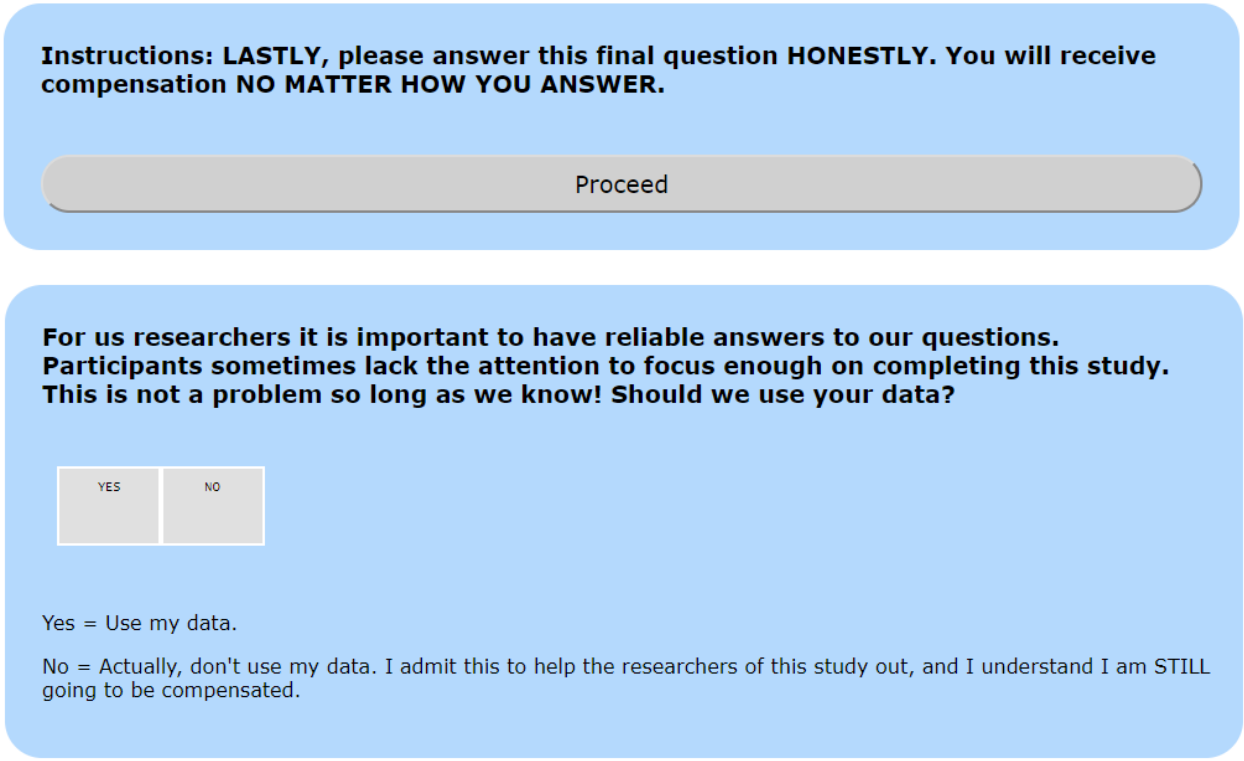
A.3 Sample Rating Task



A.4 Demographic Information



A.5 Manipulation Check

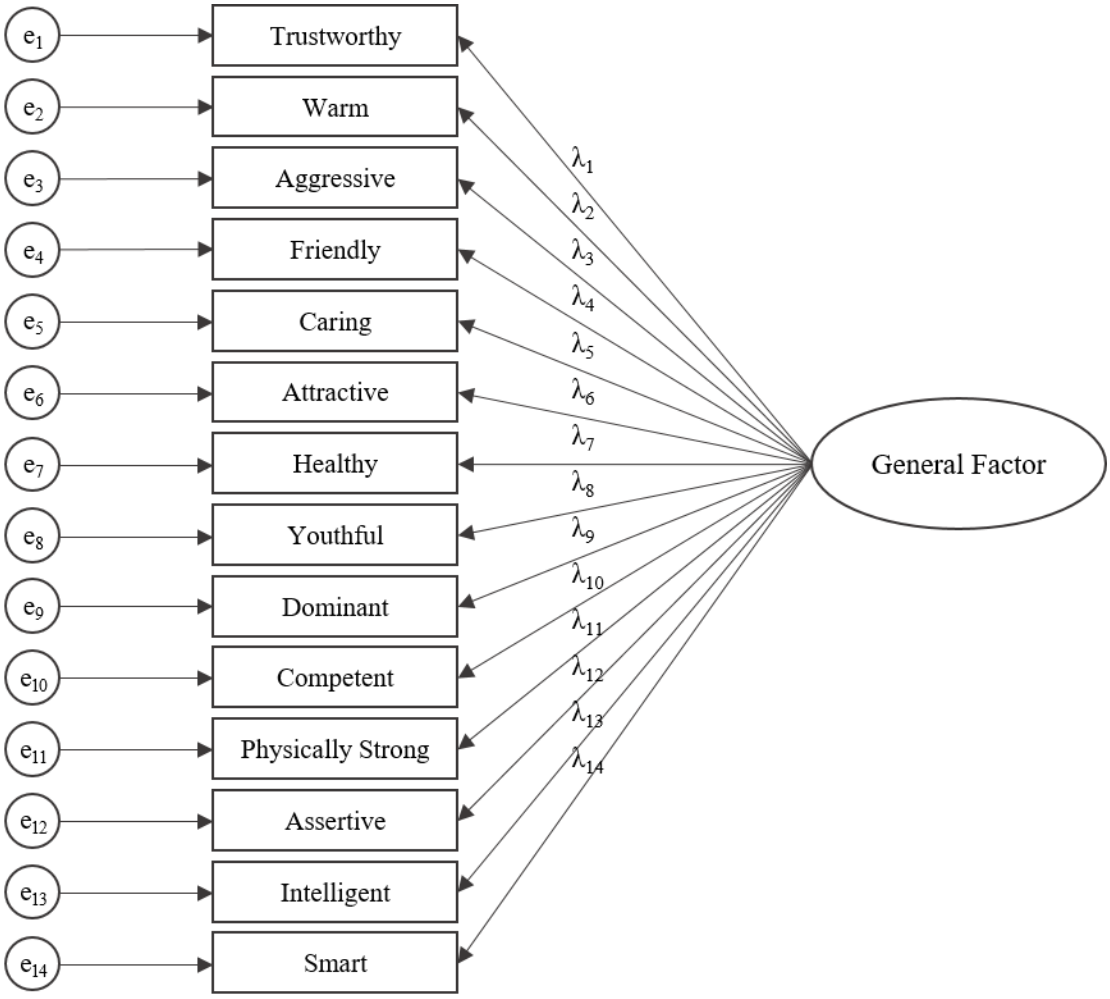


A.6 Example Stimuli

Appendix B: Supplementary Analyses

B.1 One-Factor Model in CFA

Supplementary Figure 1. SEM path diagram for a confirmatory factor analysis, representing the loadings of 14 traits onto a single general factor. For each trait, participants' ratings of each stimulus target were averaged, such that the mean rating on each trait served as the unit of analysis. e = error. λ = regression coefficient indicating the causal effect of the latent variable on the observed variable.



Supplementary Table 1

Indices of model fit for an exploratory one-factor model of face perception across own-group and cross-group perceptions.

	Chi-square ($\chi 2$)	*p* ($\chi 2$)	CFI	RMSEA	SRMR
Model					
White perceivers rating White males	1252.54	<.001	.432	.319	.255
White perceivers rating Black males	858.93	<.001	.555	.260	.178
White perceivers rating Asian males	737.42	<.001	.561	.239	.162
White perceivers rating White females	980.32	<.001	.570	.281	.198
White perceivers rating Black females	847.37	<.001	.657	.263	.144
White perceivers rating Asian females	746.81	<.001	.574	.260	.194

Note. CFI = Comparative fit index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean squared residual.

Below are general rules for acceptable fit if data are continuous (Hu & Bentler, 1999; Schreiber et al., 2006):

1. Chi-square: Ratio of $\chi 2$ to *df* ≤ *2 or 3*.
2. CFI: ≥ .95 for acceptance.
3. RMSEA: < .06 to .08.
4. SRMR: ≤ .08

Supplementary Table 2

Parameter estimates for the unidimensional model of White participants' perceptions of female and male White, Black, and Asian faces.

	White Male		Black Male		Asian Male		White Female		Black Female		Asian Female	
Factor Loadings	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
One-factor model												
Caring	1.000	---	1.000	---	.908	.067	.721	.070	.852	.036	.903	.047
Warmth	.975	.050	.996	.063	1.000	---	.735	.069	1.000	---	.988	.047
Friendly	.973	.050	.879	.062	.960	.065	.717	.069	.922	.039	1.000	---
Attractive	.516	.089	.558	.067	**.296**	.075	1.000	---	.674	.070	.510	.082
Trustworthy	.519	.074	.632	.051	.727	.065	.728	.066	.643	.047	.625	.053
Aggressive	-.924	.066	-.864	.069	-.846	.069	-.481	.078	-.867	.045	-.770	.055
Healthy	.679	.095	.605	.080	.463	.079	.991	.074	.722	.068	.629	.075
Youthful	.533	.143	**.346**	.123	**.206**	.140	.635	.099	.488	.077	.479	.095
Dominant	-.552	.105	-.518	.079	-.577	.079	**-.012**	.071	-.427	.049	-.406	.074
Physically Strong	**-.211**	.106	**-.024**	.092	**-.296**	.093	**.123**	.060	**-.067**	.049	**-.063**	.072
Intelligent	.641	.068	.687	.056	.627	.064	.785	.056	.702	.049	.435	.063
Smart	.581	.071	.596	.058	.589	.072	.820	.054	.612	.041	.442	.058
Competent	.539	.073	.590	.058	.539	.060	.975	.070	.675	.051	.485	.063
Assertive	**-.319**	.082	**-.238**	.080	**-.325**	.070	**.105**	.071	**-.184**	.049	**-.237**	.070

Note. In each model, the highest-loading trait on a dimension has an estimate (*B*) of 1.000 and no stand error (*SE*). General guidelines for interpreting the above estimates: Indicators that load strongly on its latent dimension have an estimate of > .800. Indicators that load moderately well have an estimate between .400-.800. Indicators that load weakly have an estimate < .400.

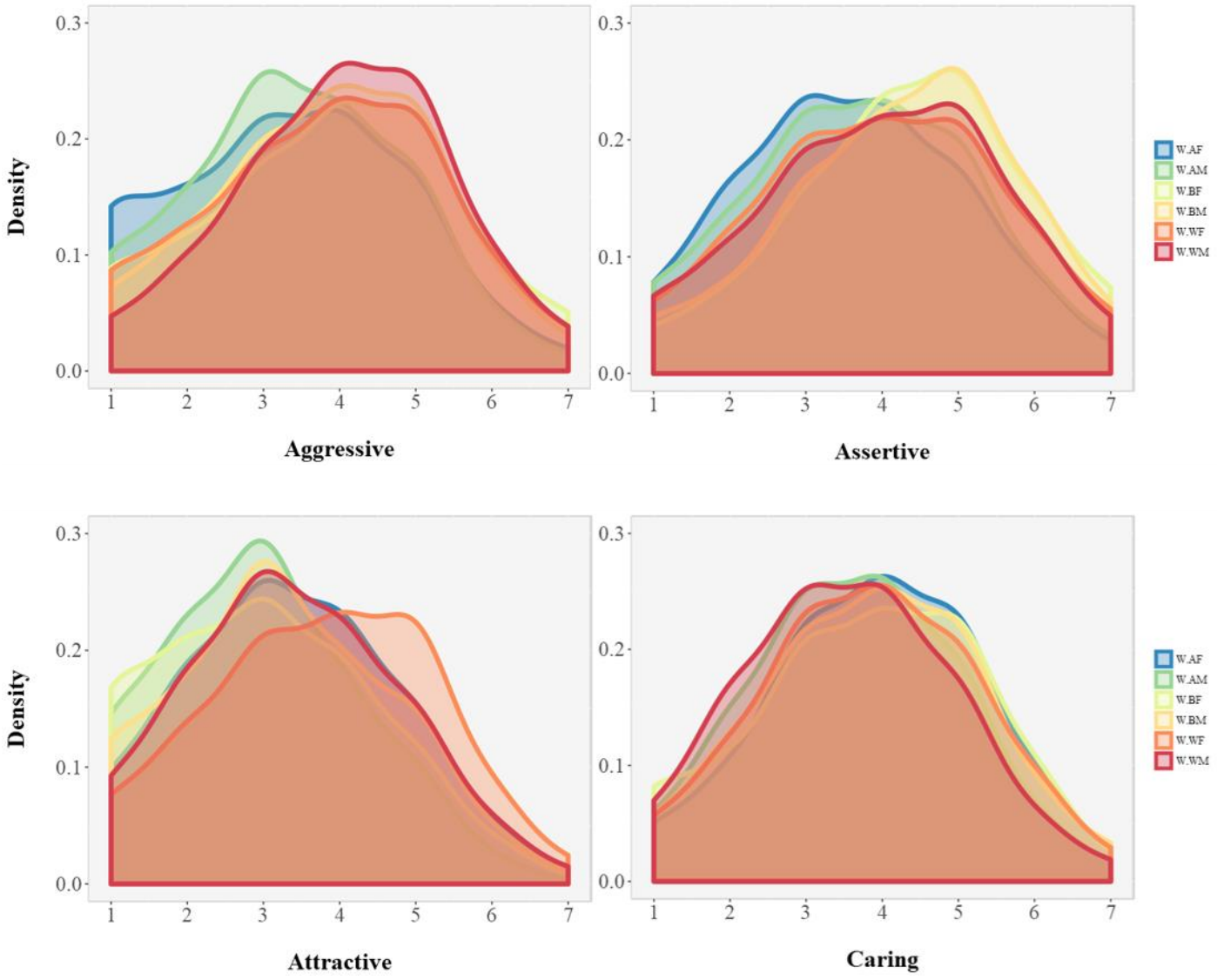
B = estimate. *SE* = stand error. *p* = p-value.

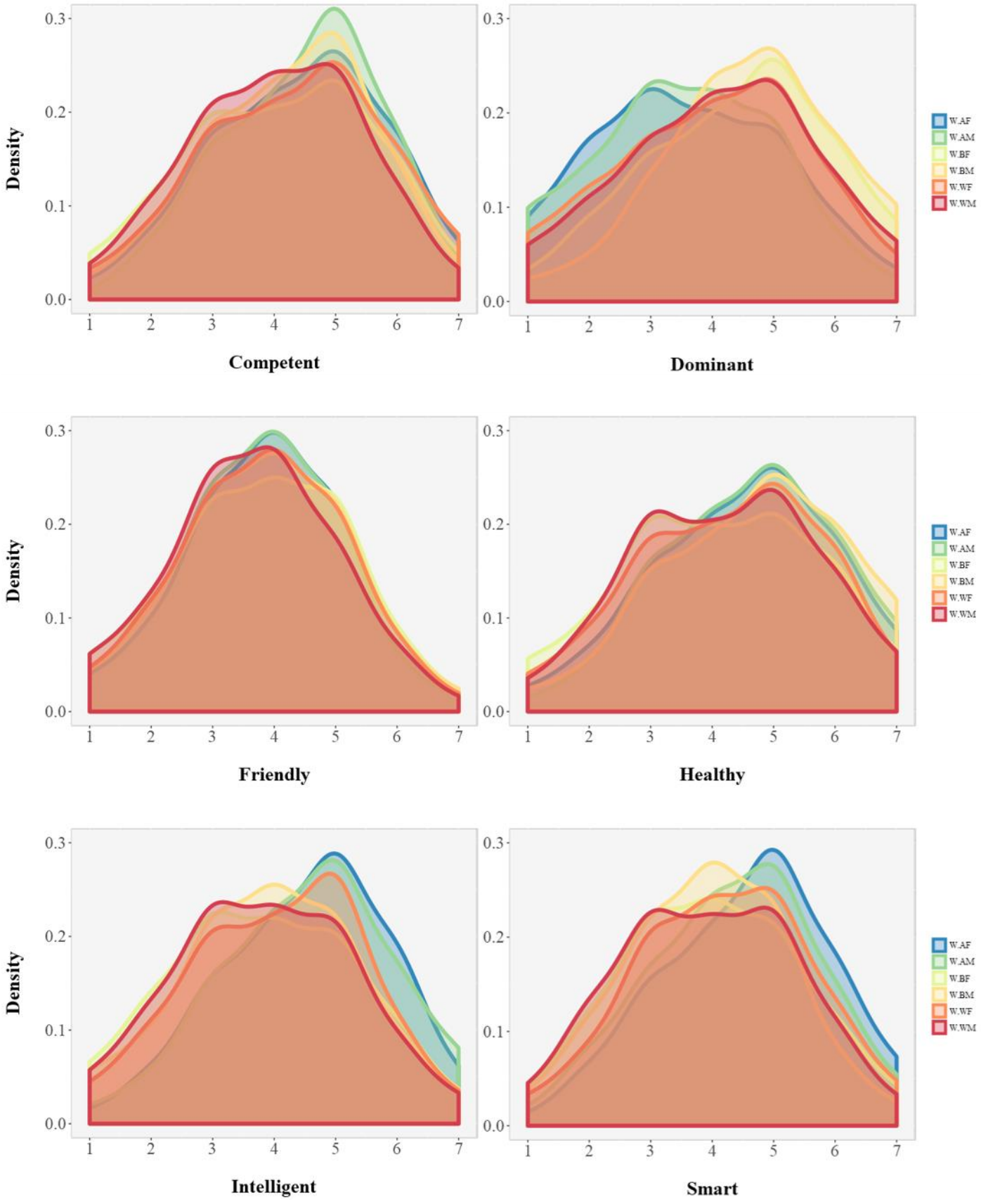
Traits which load poorly on its latent factor are **bolded**.

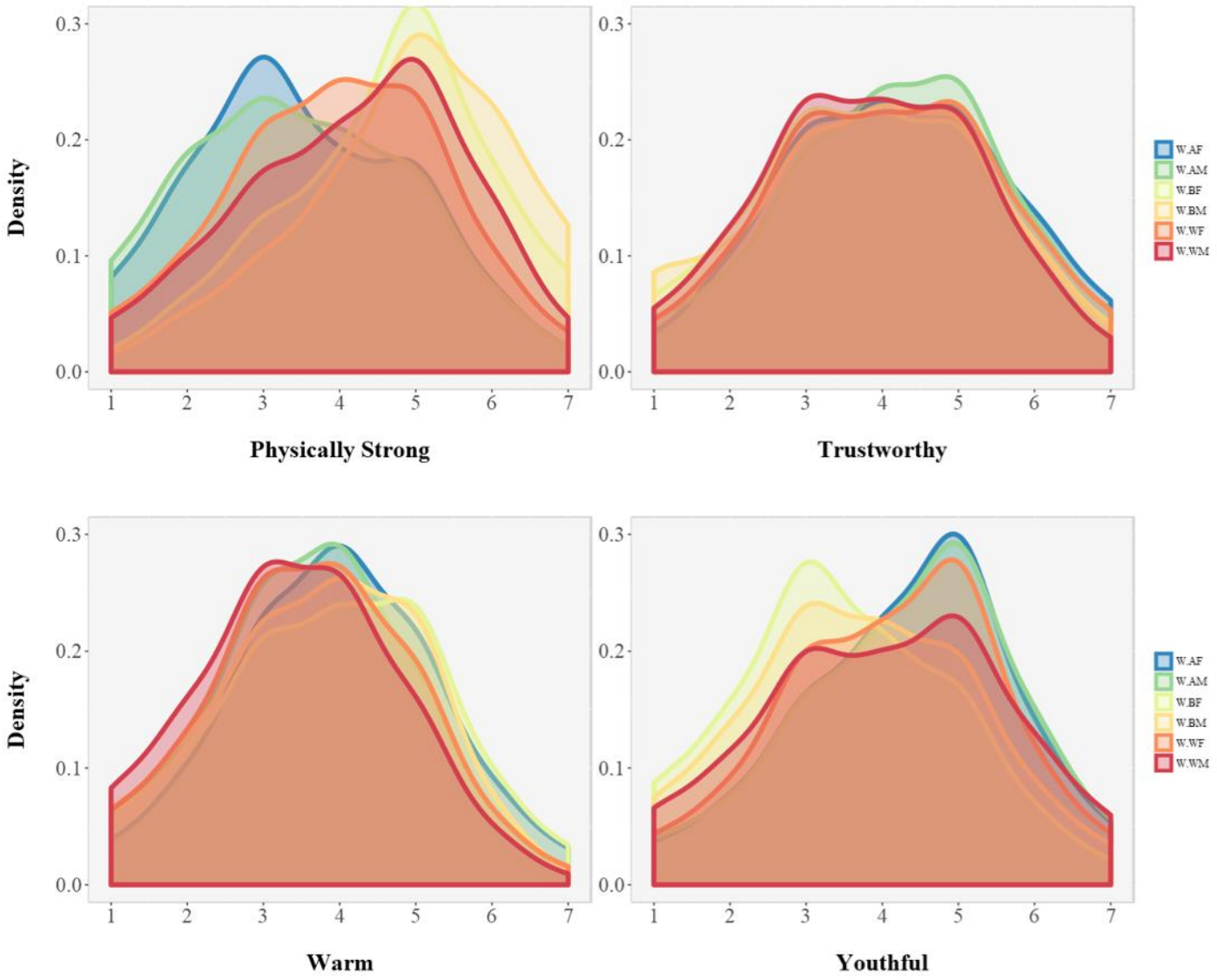
Significant (p < .05) covariances between the two latent dimensions are **bolded**.

B.2 Density Plots of Ratings on Each Trait

Supplementary Figure 2. Density plots showing the distributions of 14 trait ratings for each social group. Each distribution represents White participants' ratings of White, Black, or Asian female or male targets (e.g., *W.WF* = White perceivers' ratings of White female targets).




Density
0.0
0.1
0.2
0.3
1
2
3
4
5
6
7
Competent
Dominant
Friendly
Healthy
Intelligent
Smart
W.AF
W.AM
W.BF
W.BM
W.WF
W.WM


Density
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
Physically Strong
Trustworthy
Warm
Youthful
W.AF
W.AM
W.BF
W.BM
W.WF
W.WM

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. https://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. https://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. https://doi.org/10.1098/rstb.2015.0367

Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., Winquist, L., & Yu, D. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology*, *72*(3), 558.

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

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–53. https://doi.org/10.1073/pnas.0705435104

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706–712.

Berry, D. S., Wero, F., & Julia, L. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of Personality*, *61*(4), 497–520.

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. https://doi.org/0803973233

Blair, I. V. Y., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial

features in criminal sentencing. *Psychological Science*, *15*(10), 674–679.

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. https://doi.org/10.1145/311535.311556

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(9), 1063–1074.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*(4), 703–706.

Brewer, M. B. (1988). A dual process model of impression formation. In *Advances in Social Cognition* (Vol. 1, pp. 1–36). https://doi.org/doi:10.1521/soco.1993.11.1.150

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

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166.

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. https://doi.org/10.1111/cogs.12231

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3–25.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651.

Carpinella, C. M., Hehman, E., Freeman, J. B., & Johnson, K. L. (2015). The Gendered Face of Partisan Politics: Consequences of Facial Sex Typicality for Vote Choice. *Political Communication*, (August), 1–18. https://doi.org/10.1080/10584609.2014.958260

Carpinella, C. M., & Johnson, K. L. (2013). Politics of the face: The role of sex-typicality in trait assessments of politicians. *Social Cognition*, *31*(6), 770–779. https://doi.org/10.1521/soco.2013.31.6.770

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. https://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. https://doi.org/10.1068/p6543

Castelli, L., Carraro, L., Ghitti, C., & Pastore, M. (2009). The effects of perceived competence

and sociability on electoral outcomes. *Journal of Experimental Social Psychology*, *45*(5), 1152–1155.

Chelnokova, O., Laeng, B., Eikemo, M., Riegels, J., Løseth, G., Maurud, H., … Leknes, S. (2014). Rewards of beauty: the opioid system mediates social motivation in humans. *Molecular Psychiatry*, *19*(7), 746.

Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, *51*, 27–33.

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, *9*(3), 245–274.

Cogsdill, E. J., Todorov, A., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces a developmental study. *Psychological Science*, *25*(5), 1132–1139.

Cohen, C. E. (1981). Person categories and social perception: Testing some boundaries of the processing effects of prior knowledge. *Journal of Personality and Social Psychology*, *40*(3), 441–452. https://doi.org/10.1037/0022-3514.40.3.441

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues*, *60*(4), 701–718.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, *40*(7), 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33. https://doi.org/10.1348/014466608X314935

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, *68*(April 2016), 261–279. https://doi.org/10.1037/0022-3514.68.2.261

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

DeBruine, L. M., & Jones, B. C. (2017). Face Research Lab London Set. figshare. https://doi.org/10.6084/m9.figshare.5047666

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5.

Diedrichsen, J., & Kriegeskorte, N. (2017). *Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis*. *PLoS Computational Biology* (Vol. 13). https://doi.org/10.1371/journal.pcbi.1005508

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social*

Psychological and Personality Science, *3*(5), 562–571.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540.

Dumas, R., & Testé, B. (2006). The influence of criminal facial stereotypes on juridic judgments. *Swiss Journal of Psychology*, *65*(4), 237–244.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109.

Eagly, A. H., & Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and men? *Journal of Personality and Social Psychology*, *53*(3), 451.

Eberhardt, J. L. (n.d.). Eberhardt Face Database.

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. https://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. https://doi.org/10.1037/h0030377

Esses, V. M., Jackson, L. M., Dovidio, J. F., & Hodson, G. (2005). 'Instrumental Relations among Groups: Group Competition, Conflict and Prejudice. *On the Nature of Prejudice*, *50*, 227–243.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013.

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. https://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. https://doi.org/10.1177/0956797612461508

Fessler, D. M. T., Holbrook, C., Pollack, J. S., & Hahn-Holbrook, J. (2014). Stranger danger: Parenthood increases the envisioned bodily formidability of menacing men. *Evolution and Human Behavior*, *35*(2), 109–117. https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2013.11.004

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.

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

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. https://doi.org/10.1037//0022-3514.82.6.878

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

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*(3), 473–489.

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *The American Naturalist*, *139*(3), 603–622.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (Second). Thousand Oaks {CA}: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

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), 247–279. https://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. https://doi.org/10.1016/j.jesp.2009.10.002

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(32), 10573–10581. https://doi.org/10.1523/JNEUROSCI.5063-13.2014

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology*, *4*(1), 1–26.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491.

Graham, J. R., Harvey, C. R., & Puri, M. (2016). A corporate beauty contest. *Management Science*.

Guerra, R., Gaertner, S. L., António, R., & Deegan, M. (2015). Do we need them? When immigrant communities are perceived as indispensable to national identity or functioning of the host society. *European Journal of Social Psychology*, *45*(7), 868–879. https://doi.org/10.1002/ejsp.2153

Günaydin, G., Zayas, V., Selcuk, E., & Hazan, C. (2012). I like you but I don't know why: Objective facial resemblance to significant others influences snap judgments. *Journal of Experimental Social Psychology*, *48*(1), 350–353.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66.

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

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. https://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*, *41*(8), 1–12. https://doi.org/10.1177/0146167215591495

Hehman, E., Ingbretsen, Z. A., & Freeman, J. B. (2014). The neural basis of stereotypic impact on multiple social categorization. *NeuroImage*, *101*, 704–711. https://doi.org/10.1016/j.neuroimage.2014.07.056

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. https://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. https://doi.org/10.1016/j.jesp.2015.03.007

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

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology*, *46*(2), 445–448. https://doi.org/10.1016/j.jesp.2009.11.008

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The Unique Contributions of Perceiver and Target Characteristics in Person Perception. *Journal of Personality and Social Psychology: Attitudes and Social Cognition.*

Hehman, E., Volpert, H. I., & Simons, R. F. (2013). The N400 as an index of racial stereotype accessibility. *Social Cognitive and Affective Neuroscience*, nst018.

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, *24*(4), 265–283.

Holbrook, C., Fessler, D. M. T., & Navarrete, C. D. (2016). Looming large in others' eyes: racial stereotypes illuminate dual adaptations for representing threat versus prestige as physical size. *Evolution and Human Behavior*, *37*(1), 67–78. https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2015.08.004

Holbrook, C., Galperin, A., Fessler, D. M. T., Johnson, K. L., Bryant, G. A., & Haselton, M. G.

(2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, *14*(3), 455.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives Cutoff Criteria for Fit Indexes in Covariance Structure Analysis : Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

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

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340.

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052–1072.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168.

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology*, *44*(5), 1418–1420.

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

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is Gendered : How Covarying Phenotypes and Stereotypes Bias Sex Categorization, *102*(1), 116–131. https://doi.org/10.1037/a0025335

Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. *ETS Research Report Series*, *1970*(2).

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899.

Keating, C. F., Mazur, A., & Segall, M. H. (1981). A cross-cultural exploration of physiognomic traits of dominance and happiness. *Ethology and Sociobiology*, *2*(1), 41–48.

Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology*, *48*(1), 77–85.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165–182.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

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. https://doi.org/10.1016/j.jesp.2006.10.023

Kubota, J. T., & Ito, T. A. (2017). Rapid race perception despite individuation and accuracy goals. *Social Neuroscience*, *12*(4), 468–478. https://doi.org/10.1080/17470919.2016.1182585

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388.

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. Retrieved from https://cran.r-project.org/package=ez

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies. *World Politics*, *62*(4), 561–593. https://doi.org/DOI: 10.1017/S0043887110000195

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

Lenz, G. S., & Lawson, C. (2011). Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance. *American Journal of Political Science*, *55*(3), 574–589. https://doi.org/10.1111/j.1540-5907.2011.00511.x

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31*(1), 34–47.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, *28*(1), 18–27.

Livingston, R. W., & Pearce, N. A. (2009). The Teddy-Bear Effect Does Having a Baby Face Benefit Black Chief Executive Officers? *Psychological Science*, *20*(10), 1229–1236. https://doi.org/10.1111/j.1467-9280.2009.02431.x

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

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Macrae, C. N., & Shepherd, J. W. (1989). Do criminal stereotypes mediate juridic judgements? *British Journal of Social Psychology*, *28*(2), 189–191.

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why Do Fear and Anger Look the Way They Do? Form and Social Function in Facial Expressions. *Personality and Social Psychology Bulletin*, *31*(1), 73–86. https://doi.org/10.1177/0146167204271306

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach-and avoidance-related behaviors. *Emotion*, *5*(1), 119.

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*(5), 806–816.

Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love. *Psychological Science*, *16*(3), 236–239. https://doi.org/10.1111/j.0956-7976.2005.00809.x

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology*, *31*(1), 41–58.

McArthur, L. Z., & Apatow, K. (1984). Impressions of Baby-Faced Adults. *Social Cognition*, *2*(4), 315–342. https://doi.org/10.1521/soco.1984.2.4.315

McArthur, L. Z., & Berry, D. S. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology*, *18*(2), 165–192.

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. https://doi.org/10.3758/BF03206543

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*(4), 237–254.

Montepare, J. M., & Zebrowitz-McArthur, L. (1989). Children's perceptions of babyfaced adults. *Perceptual and Motor Skills*, *69*(2), 467–472.

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116–139. https://doi.org/10.1080/10705511.2014.961800

Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social Forces*, 823–850.

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.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A

Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003553

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. https://doi.org/http://dx.doi.org.ezproxy.snhu.edu/10.1037/0022-3514.35.4.250

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324. https://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. https://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. https://doi.org/10.1037/a0014520

Peeters, G. (2002). From good and bad to can and must: Subjective necessity of acts associated with positively and negatively valued stimuli. *European Journal of Social Psychology*, *32*(1), 125–136.

Pillemer, J., Graham, E. R., & Burke, D. M. (2014). The face says it all: CEOs, gender, and predicting corporate performance. *The Leadership Quarterly*, *25*(5), 855–864.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477–491.

Posamentier, M. T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychology Review*, *13*(3), 113–143.

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2008). Multilevel structural equation modeling. *Handbook on Structural Structural Equation Models*, 20. https://doi.org/10.1016/B978-044452044-9/50013-6

Ratner, K. G., & Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology*, *49*(2), 298–301. https://doi.org/10.1016/j.jesp.2012.10.017

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897.

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. https://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). https://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). https://doi.org/10.1371/journal.pone.0017859

Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the morality of an aggressor: the role of perceived motive. *Journal of Personality and Social Psychology*, *83*(4), 789.

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., 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.

Robinson, K., Blais, C., Duncan, J., Forget, H., & Fiset, D. (2014). The dual nature of the human face: there is a little Jekyll and a little Hyde in all of us. *Frontiers in Psychology*, *5*, 139.

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

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Ghent, Belgium: Ghent University*.

Ruble, T. L. (1983). Sex stereotypes: Issues of change in the 1970s. *Sex Roles*, *9*(3), 397–402.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105. https://doi.org/10.1016/j.jesp.2007.12.001

Rule, N. O., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, *19*(2), 109–111.

Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success. *Sex Roles*, *61*(9–10), 644–652.

Rule, N. O., Ambady, N., Adams Jr, R. B., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: prediction and consensus across cultures. *Journal of Personality and Social Psychology*, *98*(1), 1.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*, *45*(6), 1245–1251. https://doi.org/10.1016/j.jesp.2009.07.010

Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: facial maturity modulates recognition of fearful and angry expressions. *Emotion*, *9*(1), 39.

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

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*.

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1431), 1913–1917.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., King, J., Nora, A., & Barlow, E. A. (2006). Reportig Structural Equation Modeling and Confirmatory Factor Analysis Results : A Review. *The Journal of Educational Research*, *99*(6), 232–338. https://doi.org/10.3200/JOER.99.6.323-338

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. *Person Perception and Interpersonal Behavior*, 300–315.

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. https://doi.org/10.1017/CBO9781107415324.004

Secord, P. F., Dukes, W. F., & Bevan, W. (1954). Personalities in faces: I. An experiment in social perceiving. *Genetic Psychology Monographs*.

Shoemaker, D. J., South, D. R., & Lowe, J. (1972). Facial stereotypes of deviants and judgements of guilt or innocence. *Soc. F.*, *51*, 427.

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, *77*(5), 885.

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, *26*(11), 1329–1342.

Slepian, M. L., Young, S. G., Rule, N. O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition*, *30*(2), 232–240.

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. https://doi.org/10.1007/s10764-009-9380-z

Sterba, S. K., Preacher, K. J., Forehand, R., Hardcastle, E. J., Cole, D. A., & Compas, B. E. (2014). Structural Equation Modeling Approaches for Analyzing Partially Nested Data. *Multivariate Behavioral Research*, *49*(2), 93–118. https://doi.org/10.1080/00273171.2014.882253

Stevenage, S. V., & McKay, Y. (1999). Model applicants: The effect of facial appearance on recruitment decisions. *British Journal of Psychology*, *90*(May), 221–234.

https://doi.org/10.1348/000712699161369

Stolier, R. M., Hehman, E., & Freeman, J. B. (2018). A Dynamic Structure of Social Trait Space. *Trends in Cognitive Sciences*, *22*(3), 197–200. https://doi.org/10.1016/j.tics.2017.12.003

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, *48*(3), 1197–1204. Retrieved from https://web.stanford.edu/group/mcslab/cgi-bin/wordpress/examine-the-research/

Sutherland, C. A. M. (2015). First impressions from faces: Integrating facial dimensions and social categories, (March), 1–235. https://doi.org/10.1017/CBO9781107415324.004

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. https://doi.org/10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, *157*, 257–267. https://doi.org/10.1016/j.cognition.2016.09.006

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186–208. https://doi.org/10.1111/bjop.12085

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. https://doi.org/10.1111/bjop.12206

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Chicago: Nelson.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, *3*(2), 119–127. https://doi.org/10.1093/scan/nsn009

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. https://doi.org/10.1037/a0032335

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, *5*(10), 775–791.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of Competence from Faces Predict Election Outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*(15), 1–27. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, *28*(2), 117–122.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

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

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Ullman, J., & Bentler, P. (2013). Structural equation modeling. *Handbook of Psychology, Research Methods in Psychology*, 661–690. https://doi.org/10.1002/0471264385.wei0224

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161–204.

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, *111*(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Verosky, S. C., & Todorov, A. (2010a). Differential neural responses to faces physically similar to the self as a function of their valence. *NeuroImage*, *49*(2), 1690–1698.

Verosky, S. C., & Todorov, A. (2010b). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*(6), 779–785.

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. https://doi.org/10.1016/j.jesp.2013.02.004

Wang, X., & Tang, X. (2009). Face Photo-Sketch Synthesis and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*.

Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial Bias in Judgments of Physical Size and Formidability: From Size to Threat Racial Bias in Judgments of Physical Size and Formidability: From Size to Threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. https://doi.org/10.1037/pspi0000092

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*(2), 222.

Wojciszke, B. (2005). Affective concomitants of information on morality and competence. *European Psychologist*, *10*(1), 60–70.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology*, *99*(6), 964.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition a critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116–142.

Yu, W. (2002). PhD Thesis - COMPETITION AND COOPERATION IN MULTI-USER c Copyright by Wei Yu 2002 All Rights Reserved, (June).

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm... really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, *8*(3), 291–308.

Zebrowitz-McArthur, L., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238. https://doi.org/10.1037//0033-295X.90.3.215

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Westview Press.

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., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*(2), 238–249.

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. https://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. https://doi.org/10.1037/a0017990

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*(6), 603.

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

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior*, *28*(3), 167–185.