University of Nevada, Reno

Individual differences in face perception

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Neuroscience

by

Kassandra R. Lee

Dr. Michael Webster, Dissertation Advisor

May 2024

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

Kassandra R. Lee

entitled

Individual differences in face perception

be accepted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Michael Webster, Ph.D.
Advisor

Fang Jiang, Ph.D.
Committee Member

Michael A. Crognale, Ph.D.
Committee Member

Jennifer Hoy, Ph.D.
Committee Member

Dennis Mathew, Ph.D.
Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean
Graduate School

May, 2024

Abstract

Faces are ubiquitous in our daily lives, and our ability to perceive and process them drives many important tasks, such as identifying age, emotional state, and health. Despite a rich body of literature on face perception, the mechanisms involved in perceiving and interpreting faces remain enigmatic. In this work I explore individual variations in face perception. Individual differences have been studied across many domains of perception because these differences provide rich insight into perceptual mechanisms and their variability. For higher level, more complex stimuli like faces, differences themselves have been identified but the bases for them are poorly understood. Across three studies I examine individual differences in how human faces are categorized, with the goal of characterizing the nature of the variability and the underlying processes involved. In my first experiment I measured face categorization judgments for race and sex for observers living in Reno, Nevada and Tokyo, Japan, to compare both the differences between and within cultural contexts. I found large and reliable individual differences in face categorization boundaries, which were substantially larger within than between groups. In a second experiment I tested whether these categorization judgments reflect general biases (e.g., to see a face as more female) or are specific to properties of the individual faces being judged. I found biases at the level of the face categories that were equally due to differences in observers and the face identities being judged. Finally, I test the hypothesis that differences in face judgments partly reflect differences in how face coding mechanisms are normalized in individual observers. Here, I found evidence for sensitivity differences across observers, likely driven by their long-term experience of the diet of faces they encounter. By utilizing individual differences, these studies provide a rich characterization of the patterns of variation in judgments about faces – one of the most important visual stimuli for humans – and reveal insights into the bases for these differences.

Acknowledgments

I would like to acknowledge the esteemed group of individuals comprising the academic community at the University of Nevada, Reno, specifically in the Neuroscience and CBS programs. I am forever grateful for your advice, support, and dedication to my work and future success. I cannot continue without first and foremost acknowledging Mike Webster, whose unwavering patience, support, and endless expertise has motivated the direction of my research and forever altered my career path for the better. Thank you also to Fang Jiang, whose guidance helped to integrate me to UNR, and to Michael Crognale for always being available to provide advice. Thanks to Sean O'Neil for answering every text about MATLAB. I would like to express my gratitude to my cohort and other students at UNR who proved not only to be supportive colleagues, but endearing friends. This includes AR, to whom I owe most early successes and late nights finishing papers and abstracts in lab, II, CM, and KE for welcoming me. To BS, BS, and CS for being some of my first friends here. I want to thank OBK and GF for helping me keep it together during those COVID years. I want to thank MW, without whom I would have missed out on some of my best memories, like the JM comedy show and wine and cheese. I cannot express enough love for the Webster lab who got me through every professional and some personal struggles: FB, JM, CS, MKP, ZI, DJ, and of course JM and your tortured philosophical soul-glad to have been your first contact AND you still joined the lab. Thank you to my dissertation committee for all your advice and guidance: Mike Webster, Mike Crognale, Fang Jiang, Dennis Mathew, and Jennifer Hoy. My final thanks are to my family: MC and VOC, to whom I owe every success and ounce of love, past and future.

Table of Contents

List of Figures

Chapter 1

Introduction: Utilizing individual differences in the study of face perception

1.1 General introduction

The empirical process of understanding visual perception has relied on data from many observers, which ensures reliable and replicable results (Mollon, Bosten, Peterzell, & Webster, 2017). To this end, differences that can be seen across observers have often been overlooked as noise in exchange for more generalizable, averaged result data. While some data variations are not of interest and do reflect confounding random noise, typically in measurement error, there are many variations that are systematic and provide insight into real differences in observer optical, neural, and cognitive processes that mediate our visual representations of the world. These meaningful individual differences are measured using test-retest reliability or can be correlated with some measured trait, such as a clinical diagnosis (Mollon et al., 2017). These variations come with low intra-observer error and are defined by how large and stable they are (Wilmer, 2017). The ubiquity of individual differences is undeniable, with research across all levels of perception revealing variations in visual processing. Existing research has emphasized how the study of individual differences can benefit us, allowing us to test cognitive theories (Yovel, Wilmer, & Duchaine, 2014), isolate dimensions of cognitive variation (Wilmer et al, 2012;2014), and fuel the progression of science (Germine et al., 2012).

1.2 Individual differences as an applied tool

Often motivation to study individual differences has stemmed from practical considerations (Mollon et al., 2017). One advantage of studying perception at the level of the individual is to gain an understanding of who is particularly strong or weak in a perceptual skill. It is becoming more common to develop screening procedures to vet individuals for jobs that require a certain level of perceptual expertise. One example is professions that require people to view and make judgments about a variety of faces, such as border control officers. Studies have revealed a large amount of variability in face perception, and it is useful to determine where someone may fall on this spectrum to assess if they are going to perform well at their job (Duchaine & Nakayama, 2006; Russell, Duchaine, & Nakayama, 2009; Wilmer, 2017; Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016; White, Kemp, Jenkins, Matheson, & Burton, 2014). Also, of interest is utilizing the study of individuals to compare differences at the population level. There have been studies revealing differences in perceptual performance for judging the sex and age of a face (Held, 1989; Filkowski, Olsen, Duda, Wanger, & Sabatinelli, 2017; Werner, Peterzell, & Scheetz, 1990), clinical populations (Simmons, Robertson, McKay, Toal, McAleer, & Pollick, 2009), and cultural differences, which will be discussed at length later (Masuda, Ellsworth, Mesquita, Leu, Tanida, & Van de Veerdonk, 2008; Webster & MacLeod, 2011; Voegeli, Schoop, Prestat-Marquis, Rawlings, Shackelford, & Fink, 2021a,b). It is of relevance to understand the ways in which individual populations of people perceive and respond to in- and out-group populations.

Another goal in understanding individual differences is being able to tailor stimuli for observers. A commonly used example of this is correcting for luminance sensitivity for observers (Kaiser, 1988). While this doesn't have much to do with interests specifically in individual variation, it is an example of how such knowledge can control for confounds (brightness) and strengthen experimental methods (customize a stimulus to the observer). One of the biggest motivations for studying individual differences is uncovering visual mechanisms, such as deriving spectral sensitivities of cones from individual color matches (König, & Dieterici, 1886), or identifying distinct visual motion patterns of pursuit using an individual difference paradigm (Wilmer & Nakayama, 2007). These are some of the ways individual difference studies have been utilized to provide us with a richer picture of perception. As we progress these studies our goal should be to uncover underlying mechanisms driving variation and then discover ways to capture these differences to improve research and advance the field of vision science. Depending on the stimulus and mechanism of interest, different stages of these goals will be the emphasis of individual difference research.

1.3 Established individual differences

In some cases, we have utilized the study of individual differences more heavily, yielding a thorough understanding of mechanisms driving these differences so that we can shift our goal towards developing methods, analysis, and measurements that exploit them. The field of color vision and specifically colorimetry is a good example of this. Much of the variability in receptor sensitivities for color vision is accounted for by physiological differences in the eye (MacLeod & Webster, 1983; Pokorny et al., 1987; Hammond & Caruso-Avery, 2000, Asano, Fairchild, & Blonde, 2016). These have been well studied in an attempt to do away with the use of a standard observer often utilized in color experiments. Instead, models are now being built to include estimates of the individual to better capture the perceptual variation we know arises across individuals' perception of color (Asano, Fairchild, & Blonde, 2016; Lee et al., 2020). In addition to physiology, variation arises across individuals because of the environment in which they are immersed. This has been studied for color perception in terms of adaptation of our visual systems to color in the environment (Webster, Mizokami, & Webster, 2007; Webster & Mollon, 1997; Lee & Webster, 2020), and adaptation as a source of variation is relevant to other stimuli as well. Namely, there's a lot of research showing that the way we perceive aspects of faces is shaped by the diet of faces to which we are exposed in our daily lives (Webster & MacLeod, 2011; Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C. W., 2005; Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W., & Nakayama, K., 2003). The ability for our visual system to adapt to these stimuli as we move through the world raises important questions about the ways that we exploit perceptual information and encode perceptual features across all levels of complexity.

1.4 Individual differences in face processing

Face perception is an ecologically important visual task that supports many of our social percepts and interactions. While differences in face perception have been identified, from individuals who fail to recognize familiar faces (despite otherwise healthy cognitive function), to those who are experts at the task, and many variations in between (Wilmer, 2017), it is only recently that these differences have begun to be explored, and thus, the sources of this variation remain poorly understood. Ongoing research on individual differences in face perception has a different goal from that of colorimetry, as the basis for these differences is not yet well established.

The study of individual differences in face perception includes the investigation of cognitive, genetic, and environmental contributions. One major field of research touching on environmental contributions is cross-cultural research that has explored the other-race effect (ORE), in which it has been shown that we recognize faces of our own race more accurately than of another race, often explored through population sensitivity differences to face stimuli (Feingold, 1914). A similar phenomenon has been identified for faces of the opposite sex (O'Toole, Peterson, & Deffenbacher, 1996). The ORE has been well-studied and has yielded great consideration of population-level differences in face processing, such as investigation of racial attitude and interracial contact on other and own-race memory for faces (Wan, Crookes, Dawel, Pidcock, Hall, & McKone, 2017; Meissner, & Brigham, 2001; O’Toole, Deffenbacher, Valenhtin, & Abdi, 1994; O'Toole, Peterson, & Deffenbacher, 1996). The ORE supports the idea that individuals update their face-space based on the diet of faces they encounter (Valentine, 1991; Webster & MacLeod, 2011; Rhodes, Robbins, Jaquet, McKone, Jeffery, & Clifford, 2005), which helps explain how we develop such rich and variable perceptual differences. Moving forward, understanding the ORE through an individual and population-level difference lens will help us to make better decisions for selecting individuals to do face-related jobs as well as assist in boosting diversity for things like ML models that are trained with face sets of different races and with a variety of expressions, ages, etc.

Differences have been explored across many domains of face perception, including familiarity, attractiveness, and memorability (O’Toole et al., 1994; Anzures et al., 2013). Some genetic studies have demonstrated how variable the underpinning mechanisms of individual differences in face processing can be. Twin studies conducted by Wilmer and colleagues (2009, 2010a, 2010b, 2013) showed that the amount of variability accounted for by genetics vs. environment changes greatly depending on the task observers are asked to perform. This is an example of how multidimensional face processing is and how complex the study of these stimuli can be. These considerations are going to make the study of individual differences more complicated, but our findings and the ways that we apply them will be better informed and richer in the context of this multidimensionality.

1.5 Summary

While our understanding of individual differences varies across studies and stimulus types, the aim to understand, characterize, and emphasize individual differences at multiple levels is crucial to our understanding of visual perception. Individual differences can be used as a tool to better understand perception and elucidate broader scientific principles related to mechanisms underlying our daily perceptual judgments (Peterzell, 2016; de-Wit & Wagemans, 2016; Anzures et al., 2013). I aim to explore individual differences in face categorization using three studies to identify differences at the population and individual level, and then uncover potential bases driving these differences.

Chapter 2

Individual and population differences in face categories

Kassandra R. Lee[1], Koyo Nakamura[2], Yusuke Nakashima[3], Masami Y. Yamaguchi[3], Katsumi Watanabe[2], Michael A. Webster[1]

[1]Department of Psychology, University of Nevada, Reno, [2] Waseda University, [3]Chuo University

Acknowledgments

This research has been supported by EY-010834, Grant-in-Aid for Scientific Research on Innovative Areas (17H06344), Strategic Japanese-Swiss Science and Technology Programme from JSPS

2.1 Abstract

Individuals are exposed to a vastly different diet of faces depending on their social environment. These stimulus differences could shape their perception of faces by determining their individual face-space. We explored variability in face percepts by comparing face categorization judgments for the same set of face images for adults living in Tokyo, Japan and Reno, Nevada. Stimuli were morphs between four pairs of averaged faces differing in sex (female vs. male) or race (Japanese Asian or Swiss White). Observers classified different levels of the morphs according to the four categories, with the category boundary and sensitivity estimated from probit fits to the psychometric functions. For both the sex and race judgments, the overall mean boundaries differed for certain face category judgments, but individual differences within each group were substantially larger than the within-observer variability (estimated from repeated measurements). We did not observe consistent group differences in sensitivity depending on the race or sex of the observers and the stimuli. For the specific conditions of these studies, these results instead suggest that cross-cultural factors may exert relatively limited influence compared to “within-group” differences in determining an individual’s face categories.

2.2 Introduction

Faces provide us with many types of information, including someone's age, emotional state, and health. This information drives many components of our lives, such as social interactions. For this reason, the study of face perception has been extensive and has yielded a rich body of literature. It wasn't until relatively recently, however, that we began to investigate differences in the way individuals perceive faces. Figure 2.1 from Wilmer (2017) shows the trajectory of research on individual differences in face recognition, highlighting the recent acceleration of this topic in the literature (Wilmer, 2017; Wilmer, Germine, & Nakayama, 2014). Because of the ubiquity of faces and the important purposes they serve, it is worth noting that across individuals there is systematic variability in how we perceive and process them. In the absence of brain damage, there are individuals who have been shown to fail tasks involving face recognition, while individuals on the other end excel at recognizing minimally familiar faces, as well as individuals all along this spectrum (Wilmer, 2017). Individual differences in face perception have been studied for a variety of judgments, including attractiveness, recognition, and emotion (Little, Jones, & DeBruine, 2011; White & Burton, 2022; Hamann & Canli, 2004). Individual variation in face perception judgements has been identified, but less is understood when it comes to the underlying mechanisms driving observed differences.

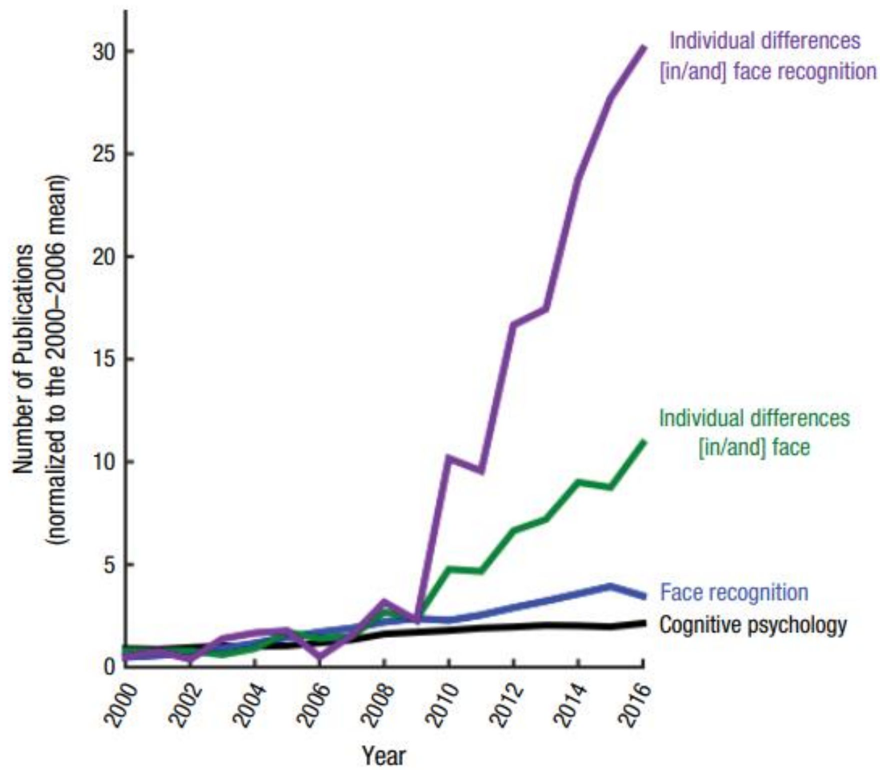


Figure 2.1. The recent acceleration of research on individual differences in face recognition. The three lines represent averaged results for searches in Google Scholar and the Web of Science Core Collection. The y-axis is the number of publications divided by the mean for 2000-2006 to equate comparison across the searches. *For details on search terms and parameters, see Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. Current Directions in Psychological Science, 26(3), 225-230.*

Because faces are variable and complex, it has been difficult to understand what drives judgments being made by individuals. There are non-clinical variations that lead to widespread differences in face perception, such as cases of individuals who are poorer than average at face recognition tasks, as well as people who are unusually good at this (Russell, Duchaine, & Nakayama, 2009; White & Burton, 2022). In an attempt to quantify differences, researchers have investigated the relationship between cognitive abilities and face recognition. Many studies have shown that variability in face recognition abilities is minimally related to/accounted for by IQ (Gignac, Shankaralingam, Walker, & Kilpatrick, 2016; Van Gulick, McGugin, & Gauthier, 2016; Wilmer, 2017). When it comes to attractiveness judgments, it has been reported that many evolutionary factors play a role in variability. This includes factors tied to individual physiology, such as hormone levels and fertility (Rhodes, 2006; Little, Jones, & DeBruine, 2011). These factors may drive judgments of attractiveness across sex in order to define preferences surrounding evolutionary success. Additional factors when making these judgments have been shown to include judgment of one's own attractiveness, as well as environmental/social factors linked to threat (Calder, Ewbank, & Passamonti, 2011). Important work by Wilmer and colleagues (2009, 2010a, 2010b) utilized twin studies to identify a genetic basis for face memory and recognition. This research showcased a phenomenon whereby the specific cognitive ability of face memory was linked strongly to heritability. Interestingly, additional twin research by Wilmer et al. (2013) showed that despite the genetic underpinnings of face recognition, judgments made on attractiveness of faces are highly affected by environmental factors. This indicates that our decisions on facial attractiveness are highly susceptible to experience. The variability in the bases for judging facial memory vs. attractiveness further highlights the complexity of the mechanisms underlying visual processing of faces. These findings are not unlike some of those identified in the underlying causes of variation in color perception. Individual differences in color can reflect sensitivity differences of an individual, but differences can also be studied separately in terms of color appearance, and the two are not predictive of one another, indicating that they depend on different factors. The ways that we perceive and categorize faces may be similar to the large, but poorly understood differences in color hue categories. If this is the case, then it is worth investigating if parallels can be drawn across judgments of these face categories to those reported for color categories (Berlin and Kay, 1969).

Differences in face judgments have been linked both to internal factors of an individual and to external factors, such as environment. One of the most well-known examples of individual differences in face perception is the other-race effect (ORE), classified by a deficit in processing faces of another race compared to our own. The primary explanation for this effect is that we are primed by the diet of faces we encounter daily and may have limited contact to individuals of other ethnicities, reducing our perceptual expertise to them (Hancock & Rhodes, Tanaka et al, 2004). ORE studies have utilized the study of individual differences to better understand the processes underlying face perception. For example, DeGutis et al. (2013) demonstrated that individuals more actively engage in a form of holistic processing for faces of their own race compared to other ethnicities. Additionally, cognitive and behavioral measures have provided evidence that we process familiar vs. unfamiliar face stimuli differently (Johnston & Edmonds, 2009). While the ORE demonstrates group level individual differences in face processing, there has also been variability identified in how much individuals are affected by it. Some cases of the ORE have proven to be so severe it mimics face-blindness for other-race faces (Wan, Crookes, Dawel, Pidcock, Hall, & McKone, 2017). This severity for a small number of individuals is mediated by factors we know drive differences across observers already, namely lack of contact with other race faces and lower than normal general face recognition abilities. While there are a small number of individuals who are so severely impacted by the ORE, it emphasizes the importance of considering the impacts of the

ORE at the individual level for specific tasks, such as eyewitness testimony or passport checking.

These studies further emphasize how the complexities of face processing can lead to rich patterns of individual differences. A review by Yovel, Wilmer, and Duchaine (2014) noted that studying these individual differences in face processing provides a powerful tool to help answer fundamental questions linked to behavioral and neural mechanisms. However, many of the individual differences in face processing remain poorly characterized, largely because faces are so complex, and we perform many perceptual tasks related to them. A more thorough look at differences in how individuals define face categories could reveal important information about face processing. Here, I will use methods designed to compare cultural differences and provide insight into the ways in which individuals categorize faces.

2.3 Experimental methods

2.3.1 Participants

Participants included 23 observers from Reno, Nevada (14 female, 1 Asian) and 13 observers from Tokyo, Japan (7 female). All participation was with informed consent and followed protocols approved by the University of Nevada, Reno IRB.

2.3.2 Stimuli

Morphs of 10 frontal-view faces were used to create averages for each of 4 categories (Asian male-White male (AMWM), Asian female- White female (AFWF), Asian female- Asian male (AFAM), White female- White male (WFWM)) across 2 dimensions of face judgments (sex, race) (Figure 2.2). The faces were obtained as still images from a Swiss-Japan facial movie database, developed by our Japan collaborators (Namba, Sato, Nakamura, & Watanabe, 2022). Note, all the Asian faces for this experiment were Japanese and all of the white faces were Swiss. The averaged faces were then morphed between different pairs to form a gradient of 100 steps within each category.

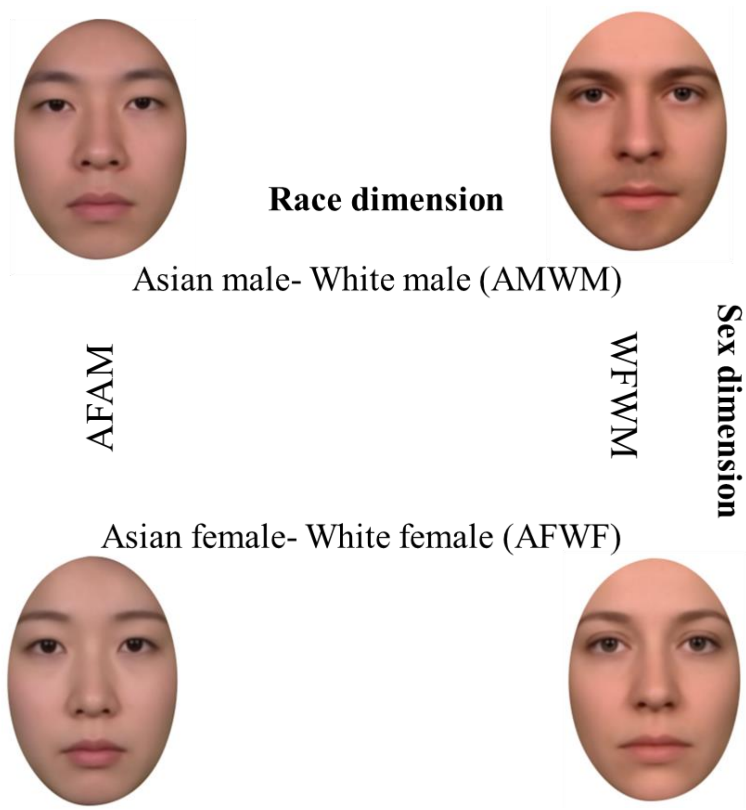


Figure 2.2. Average identities of faces used to create morphs. The four faces are the average faces representing Asian male (top left), white male (top right), Asian female (bottom left), and white female (bottom right). Pairs of these faces were then morphed along the four dimensions to create the series of faces varying in either race or sex.

2.3.3 Procedure

Stimuli were presented using Visual Basic on a 23” Dell flat panel monitor with 1920x1080 resolution and a 60Hz refresh rate. The monitor screen was calibrated with a

Photo Research PR 655 spectroradiometer. Observers viewed the display binocularly, seated 60cm away from the monitor, so stimuli subtended ~5.2° of the visual angle. For testing, 11 morph levels were shown in random order. These ranged from 25-75% in 5% steps. Each face was presented onscreen 40 times over 2 sessions for the Reno observers and 1 session for the Tokyo observers. Observers were asked to classify the face as Japanese male, Japanese female, Swiss male, or Swiss female with a key press.

2.4 Results

Reno observers completed two runs of the face classification experiment, while Tokyo observers completed one run. The boundaries estimated from the two runs across all face sets for Reno observers were highly correlated, r = 0.88, and run one correlated highly with run two for all of the categories individually, indicating that observers were consistent in their categorization across runs (relative to the within observer variability). Specifically, the difference in variance between the observers was 2.6 times greater than the variability in the repeated settings of a single observer.

2.4.1 Category boundaries

The 25 faces at each extreme end were not used for the experiment, so 11 morph levels, ranging from 25-75% in 5% steps were shown. These are represented as values -20:20 on the x-axis of Figure 2.3. Category boundaries and sensitivity were estimated for each observer using probit fits (cumulative Gaussian fit to the response proportions) to the settings of 2 observer runs for Reno observers and single run data for Tokyo observers (Figure 2.3). Here, the 50% point is the category boundary while the steepness of the curve gives a measure of sensitivity (Bliss, 1934; Cramer, 2002). While we may expect less sensitivity for out-group race judgements, sensitivity did not differ from Reno to Japan observers for any face set or by observer sex. Despite some cross-cultural differences, the variability within each test population was substantially larger; Standard deviation across observers' categories were on average 2.6x greater than the SD of the within-observer settings.

When comparing variance in boundary responses across cultures, there was significantly more variance in the Reno (M = 7.5, sd = 13.3) compared to Tokyo (M = 15.5, sd = 7.9) responses when judging Asian female to Asian male morphs, $F(22) = 2.8$, $p = .03$. Overall, there was a trend of more variance when judging the sex of the other race for both Reno and Tokyo observers, including trending significance for the WFWM category, p=.05. (Figure 2.4). The race categories did not show significantly different amounts of variance across Reno and Tokyo.

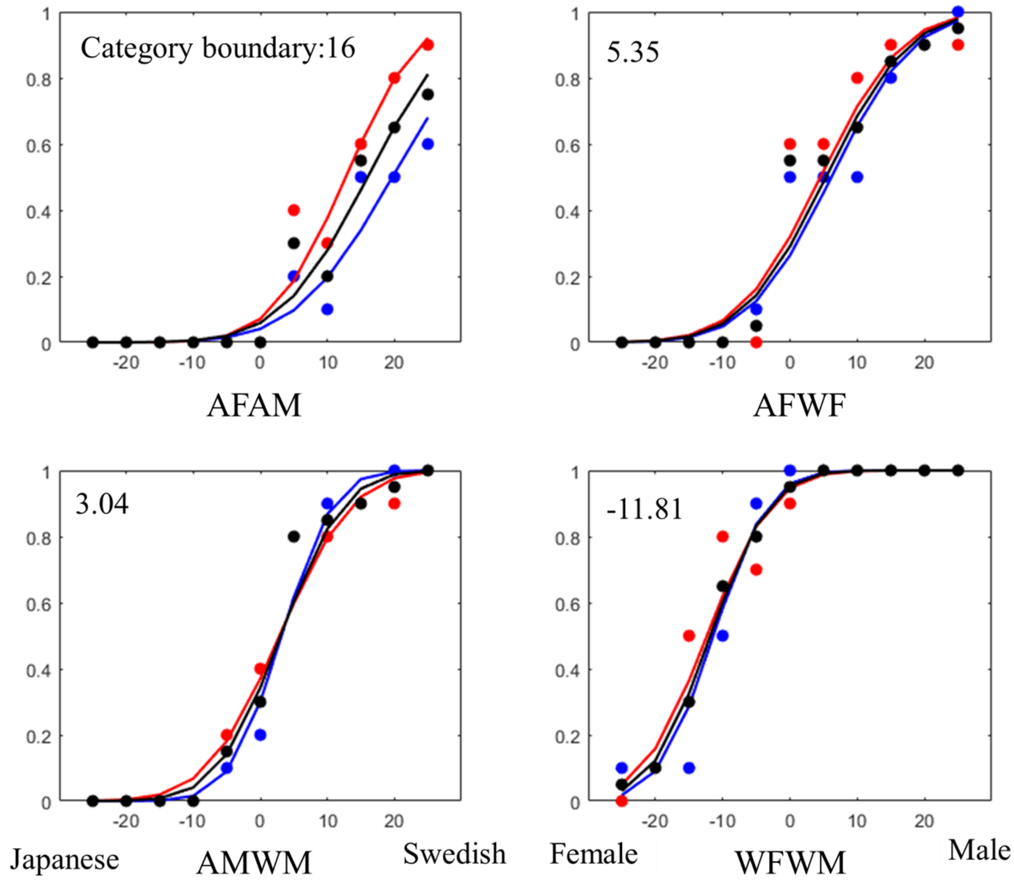


Figure 2.3. Example Probit fits for one representative observer. Category boundary (values in top left of each graph) and sensitivity (slope of lines) were estimated from probit fits to the psychometric functions. Points represent the percentages for each response while lines are the fits. Blue is run one, red is run two, and black is the mean of both.

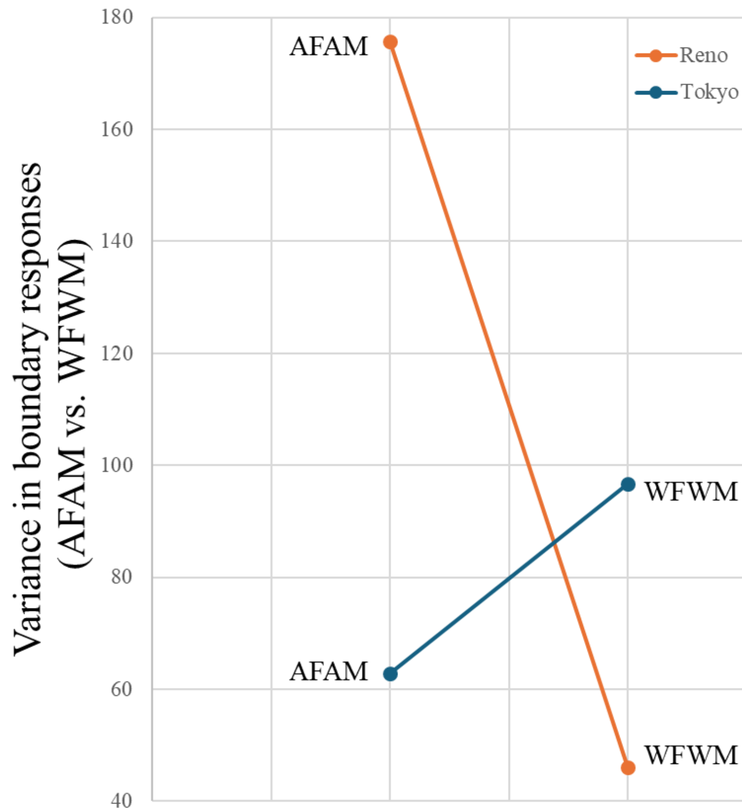


Figure 2.4. Variance in judging sex of other race faces. Differences in observer variability of boundary judgements from both Reno (orange) and Tokyo (blue) observers. Both groups of observers show higher amounts of variability in responses when judging the sex of faces of their opposite race as opposed to their in-group race.

The category boundaries of all observers for each of the face categories are shown in Figure 2.5. Each point represents an observer's boundary within that face category, with black triangles showing the average response for both Reno and Tokyo groups. We see cross-cultural differences, more emphasized in some face categories. There was a significant difference in the average categorization boundary for Reno (M = 4.26, SD =6.6) and Tokyo (M = 15.5, SD = 7.9) observers for face set AFAM, $t(34) = -4.54$, $p <$ .00. There was also a significant difference in the average categorization boundary for Reno (M = -6.4, SD =3.39) and Tokyo (M = -11.9, SD = 9.8) observers for face set

WFWM, $t(34) = 2.48$, p <=.02, and a Reno ($M = 0.30$, $SD = 3.79$) v. Tokyo ($M = 4.5$, $SD = 8.12$) difference for face set AFWF, $t(34) = -2.43$, $p = .02$. Figure 2.6 shows the same differences in cultural responses for the four face categories in a bar graph. Despite significant differences in some of these categories, the variance within a location was at least 2.6x larger in each category than the variance across locations.

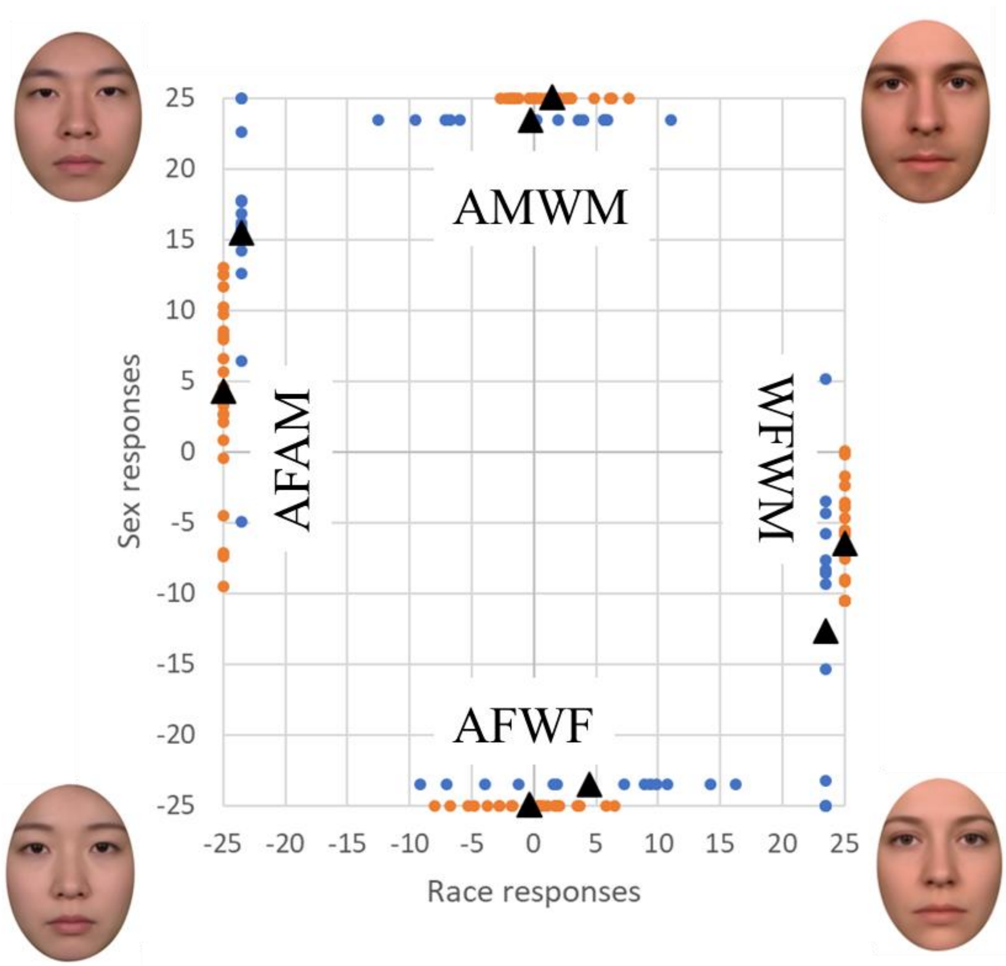


Figure 2.5. Category boundaries for all observers. Data for both Reno (orange) and Tokyo (blue) observers, showing category boundary responses for all four face categories. Each dot represents categorization data from one observer, and black triangles are average response across all subjects in that location.

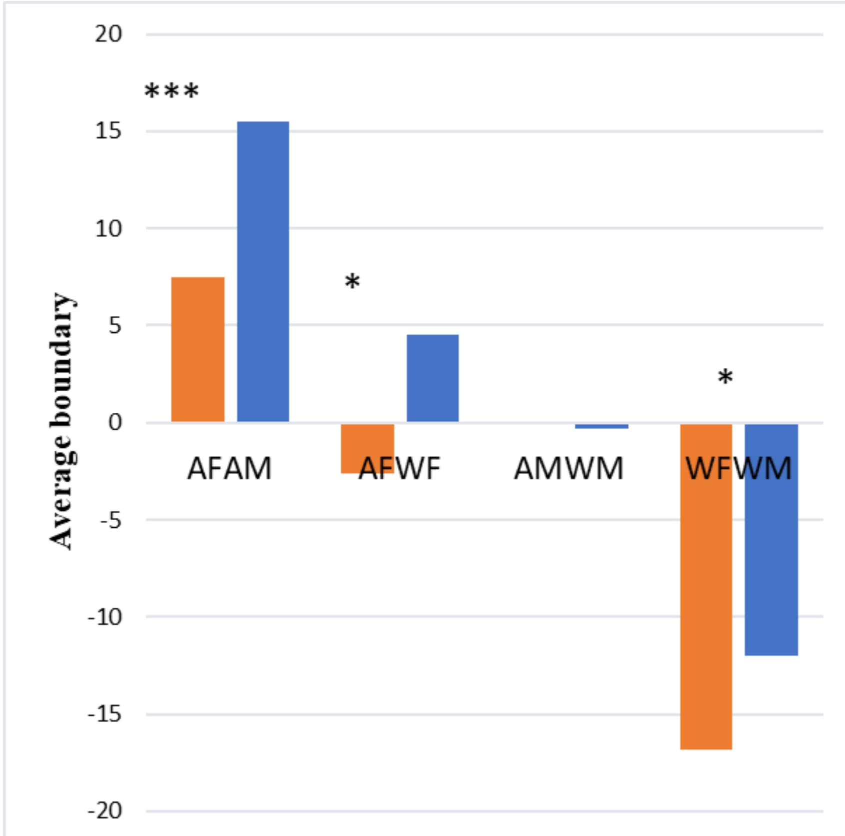


Figure 2.6. Average category boundaries. Reno (orange) and Tokyo (blue) observer category boundaries. Asterisks indicate significant differences in the group means.

2.5 Discussion

The main aim of the current study was to explore cross-cultural differences in the way people categorize faces by race and sex. We conducted a simple face categorization task in Reno, Nevada and Tokyo, Japan to explore differences between the two groups. From the categorization responses, we calculated individual and group level category boundaries to represent where each observer and cultural group fell for their categorization for each of the four face groups.

2.5.1 Cross-cultural differences

Cross cultural differences in categorizing faces may reflect a variant of the other race effect. The ORE is a widely studied phenomenon and has been shown to bias sensitivity in face perception for individuals living in primarily mono-racial societies (Anzures et al., 2013). It has been demonstrated for many race combinations, and like the current study, consistent differences have been found for Asians looking at Caucasian faces and vice versa (Hayward, Rhodes, & Schwaninger, 2008; Tullis, Benjamin, & Liu, 2014; Wan, Crookes, Reynolds, Irons, & McKone, 2015). While the ORE is typically shown to affect observer sensitivity to faces, we did not see sensitivity differences here, but instead saw a difference in the category boundaries of observers. This means observers in our Tokyo group likely showed a bias in their category boundaries based on their exposure to Japanese faces more than Swiss ones, and vice versa for the Reno observers. Importantly, if these boundaries are shifted then we might expect people to be more sensitive to how a face differs from their own race category, as was found for adaptation to natural face categories in Webster, Kaping, Mizokami, and Duhamel (2004). But that finding was not replicated here. While some studies have shown the ORE to have a modest effect (Brigham et al., 1982; Meissner & Brigham, 2001), Wan et al. (2017) reported that for some individuals the effect can cause major functional consequences for real world tasks involving faces. These population differences are thought to arise based on the ways individuals update their face-space during exposure to faces, which supports our second and more surprising finding.

2.5.2 Within-cultural differences

In addition to cross-cultural differences, we found large and reliable variation in category boundaries within our two test populations. When we consider the concept of individualized face-spaces, we can better understand where this variability may stem from. As we are exposed to faces in our environment, we update our ‘norm’ face by which we judge all other faces (Valentine, 1991; Webster & MacLeod, 2011; Rhodes, Robbins, Jaquet, McKone, Jeffery, & Clifford, 2005). If each person starts with a different norm based on their experiences, then it is very unlikely that any two individuals would have the same reference from which they perceive other faces. While this may explain the variable percepts we observed within our groups, it does seem interesting and surprising that within cultural differences would be larger than cross-cultural ones.

However, this result is similar to a pattern identified in the field of color science, where there are strong cross-linguistic similarities in color naming yet large individual differences within languages (Berlin & Kay, 1969; Lindsey & Brown, 2009). The basis for the individual differences in color appearance and color categories are not fully understood but are not the result of differences in spectral sensitivity (Emery and Webster, 2019). Instead, it is possible that the association between linguistic categories and stimuli is weak and thus can vary widely between individuals; or as noted above, that individual differences in the visual diet are large and thus lead to substantial variation. Whatever the basis, our results point to a strong similarity between color and faces in terms of how the categories vary within and across populations.

Chapter 3

Contribution of stimulus and observer differences to face categorization

Kassandra R. Lee, Michael A. Webster

Department of Psychology, University of Nevada, Reno

Acknowledgments

This research has been supported by EY-010834

3.1 Abstract

Individuals vary widely both in their ability to recognize faces and in how they classify faces along dimensions such as race, sex, or expression. However, the bases for these differences are poorly understood. In particular, it is not known how the categorization judgments depend on the properties of the specific faces. We examined the pattern of differences for typical categorical judgments of the sex or race for a wide variety of individual faces. The face images were from the Chicago Face Database for individuals labeled as male or female and Asian or White, with 10 faces selected from each of the four categories. Faces were cropped to remove external features, and then paired and morphed to form 40 stimulus sets across dimensions. The morphed faces for each set were shown simultaneously as a graded series arranged in a circle on a display and spanned the two original identities in steps of 5%. Observers selected the face closest to the category boundary for each set. Reliability for repeated measurements was high ($r \sim 0.9$), with average between-observers variance 2.3x the within-observers level. Correlations were generally weak across stimuli showing that the observers did not merely differ in a general bias but rather in face-specific biases. These results suggest that observers may partly use distinct and idiosyncratic strategies for making basic categorical distinctions between faces.

3.2 Introduction

The ability for people to complete tasks related to face identity is relevant for many real-world situations, such as eyewitness identification. Studies have shown that there are individual differences in the ways that people perceive and recognize face identity, even within expert groups. For example, in a study conducted with individuals who had specialist experience at judging faces, performance on a face comparison task was variable across viewers, with some performing well and others performing poorly (White, Kemp, Jenkins, Matheson, & Burton, 2014). This variability is strong and persistent, leading researchers to study face perception for individuals working in groups instead of alone to decrease variability and increase performance accuracy (Phillips et al., 2018; Bruce, Bindemann, & Lander, 2018; Balsdon, Summersby, Kemp, & White, 2018). Not only are we interested in studying how much variability exists across people, but we want to know what mechanisms are driving it and what this can tell us about perceptual and cognitive organization of face processing.

Such questions have begun to be answered by discovering the things that do not seem to be driving systematic differences in face processing abilities. One study aimed to understand the relationship between perceptual variability and aspects of face-identity processing. While different face tasks correlated with each other, it was found that task-specific influences drive performance and limit the amount of variability that can be accounted for with a general face-perception factor (McCaffery, Robertson, Young, and Burton, 2018). This study also ruled out that variation in perceptual abilities can be strongly accounted for by memory differences. Experience from years in a professional job that requires face matching also does not drive face matching performance (Bruce, Bindemann, & Lander, 2018). Additionally, it was found that training for face tasks does not increase our ability to complete memory and matching tasks and may in fact put us at even more of a disadvantage due to forcing unnatural processing (Woodhead, Baddeley, & Simmonds, 1978). In contrast, later work found that training individuals to detect morphed faces-a task which yields substantial individual differences-leads to higher performance (Robertson, Mungall, Watson, Wade, Nightingale, & Butler, 2018). These findings beg the need to better understand the bases underlying individual differences in face processing, and specifically for different face tasks and identities.

In a previous study (chapter 2) we found that a face categorization task yielded large and reliable individual differences across observers. A limitation of this study is that single identities were judged within each of the race and sex categories. It is of interest to investigate individual difference patterns for multiple identities within a category to see if they depend on the specific faces judged or general individual biases.

3.3 Experimental methods

3.3.1 Participants

Participants included 21 observers from Reno, Nevada, although 16 total were analyzed (8 female, 13 white, 3 Asian). Five observers were eliminated from analyses due to missing data or low reliability across runs. All participation was with informed consent and followed protocols approved by the University of Nevada, Reno IRB.

3.3.2 Stimuli

Face images are from the Chicago Face Database for individuals labeled as male or female and Asian or White, with 10 faces selected from each of the 4 categories (Ma, Correll, & Wittenbrink, 2015). These were cropped to remove external features, and then paired and morphed to form 40 stimulus sets (Figure 3.1). The faces that were paired to form each morph were randomly selected if they were classified by the database as White, Asian, Female, and Male. All faces were morphed by one experimenter for consistency and then checked by a second experimenter. If any of the morphs looked like the identities did not align well (e.g., blur around the eyes because they were drastically different in shape or size) then those morphs were redone with new identities. Face categories are abbreviated AFAM: Asian female-Asian male, WFWM: White female-White male, AFWF: Asian female-White female, and AMWM: Asian male, White male.

Figure 3.1. Examples of face identities. Examples of eight individual faces used to morph across sex and race dimensions. *Left panel* shows examples in the sex dimension: AFAM, WFWM. *Right panel* shows examples in the race dimension: AFWF, AMWM. For each identity pair the extreme identities are on the outside, while the 50% morph is in the middle. Within each of the four categories, a total of 10 different morphed pairs were created; Each example here shows one.

3.3.3 Procedure

Stimuli were presented in Matlab 2022b on a 32” Cambridge Research Systems Display++ with 1920x1080 resolution and a 120Hz refresh rate. The monitor screen was calibrated with a Photo Research PR 655 spectroradiometer. Observers viewed the display binocularly, seated 80cm away from the monitor, so each individual face stimulus subtended ~4° of the visual angle. The morphed faces for a set were shown simultaneously as a graded series arranged in a circle on a display and spanned the two original identities in steps of 0.05 (Figure 3.2). Observers selected the face closest to the category boundary for each set, with the individual pairs for each dimension shown in random order, and with the morph configuration rotated randomly each trial (so that the position of the specific faces varied). Instructions specified that the observer should “select the face that is an equal balance of male/female OR White/Asian,” depending on the category being judged. All observers completed two runs of the experiment, selecting the category boundary for a given morph five times per run.

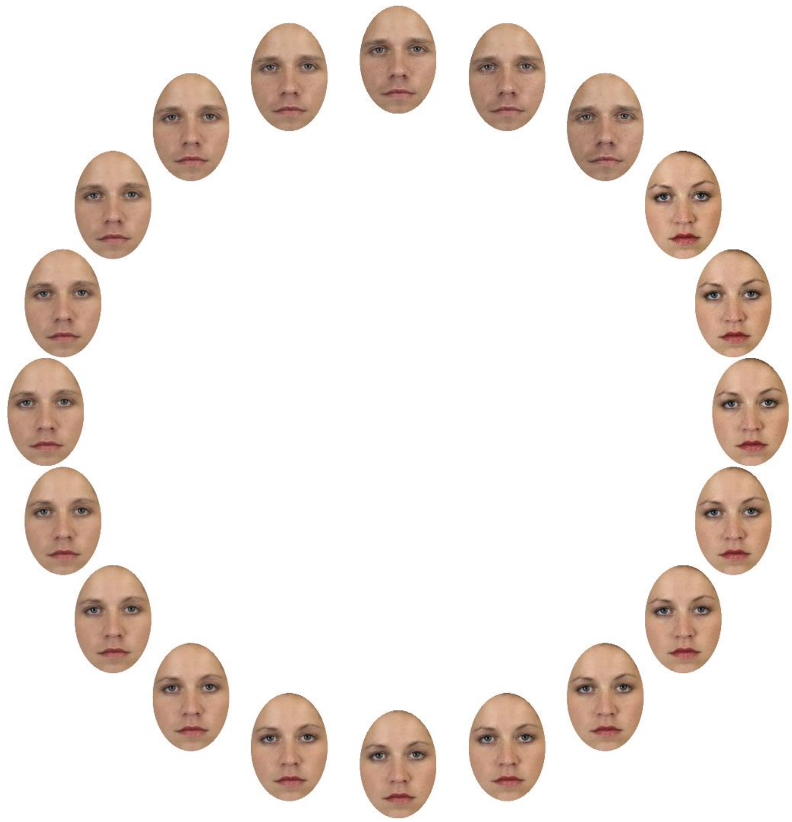


Figure 3.2. Example of the display for the sex judgment set. Faces were arranged in a circle and remained on the screen while observers selected the face that appeared to be an equal balance of male/female OR White/Asian, depending on the category. On different trials they rotated by random angles so that the morph levels were not tied to a specific angle or screen location.

3.4 Results

To assess reliability, we compared the settings across the two runs of observers. Between observer variability was 2.3x greater than within observer variability (Figure 3.3). Additionally, correlations between the two runs for each category were high (AFAM $r = 0.8$, WFWM $r = 0.5$, AFWF $r = 0.8$, AMWM $r = 0.8$), indicating observers were consistent in their boundary settings across runs. Correlating all four of the face categories showed us that people were consistent in judging race independent of sex (two race categories, $r = 0.8$), but not vice versa (two sex categories, $r = 0.4$).

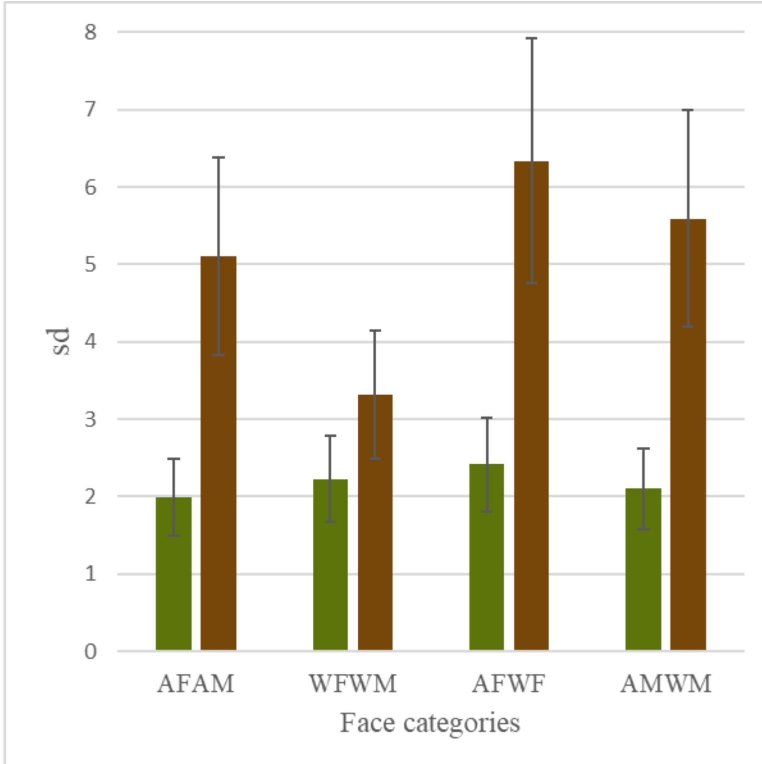


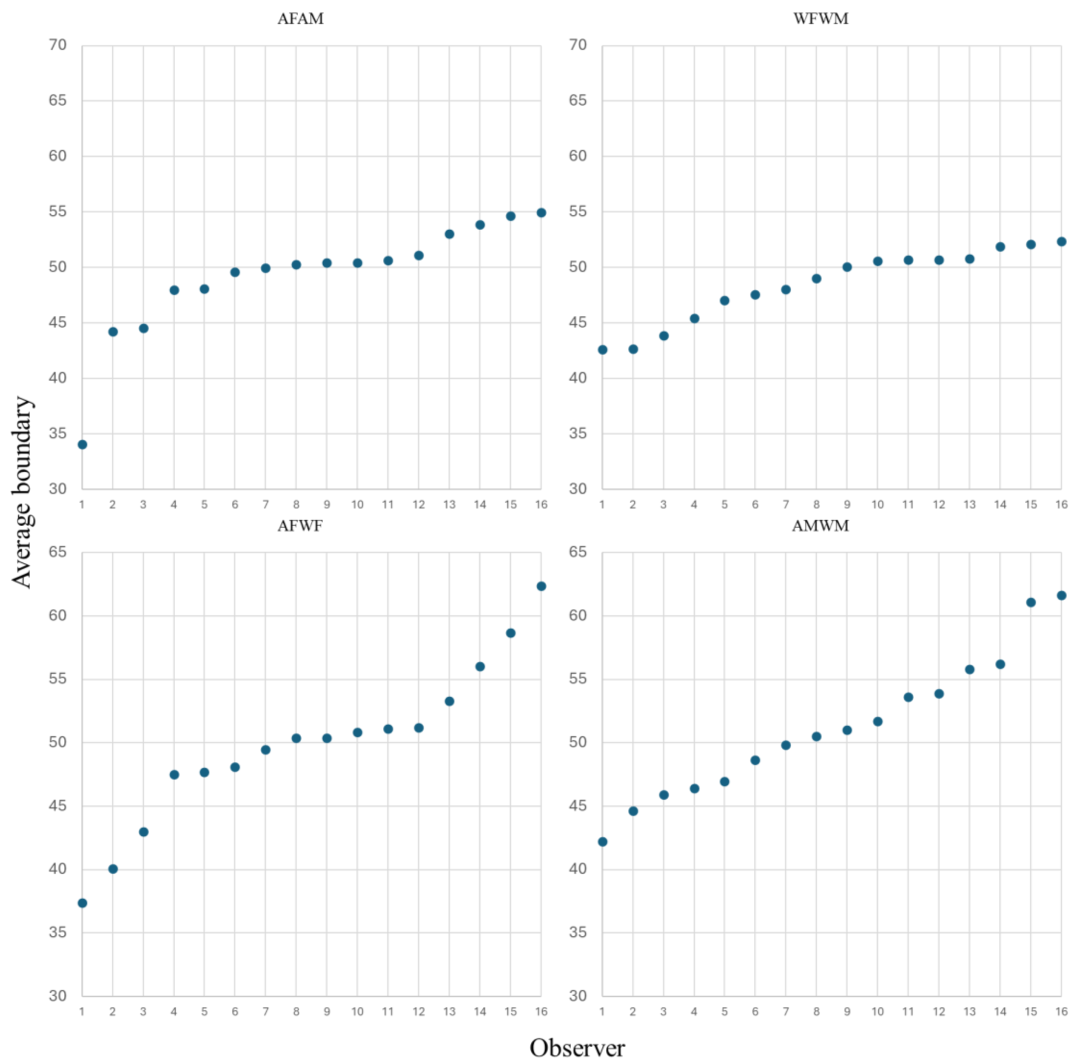
Figure 3.3. Within vs. between variability across observers. X-axis is the face categories, abbreviated: AFAM: Asian female-Asian male, WFWM: White female-White male, AFWF: Asian female-White female, and AMWM: Asian male, White male. Y-axis is SD calculated from the difference between observers' two runs. Within (green) variability is lower for all face categories than between (brown) variability, indicating that observers were more consistent within their own responses than compared to other observer responses.

3.4.1 Effect of observers and stimuli on boundaries

Individual differences we observed in boundary selection could be due to both the observers and stimuli, or perhaps more one than the other. To assess this, the variance across observers for each face set (Figure 3.4a) was compared to the variance across the ten face pair identities within each set (Figure 3.4b). The variance observed for each of these was comparable, indicating that variance we observe is due to both the observers and the different stimuli. When assessing variance across observers, there was a trend nearing significance that showed larger variance when observers judged Asian faces ($M =$

49.3, $sd = 5.1$)) compared to white faces ($M = 48.5$, $sd = 3.3$) in the sex dimension, $F(15) = 2.4$, $p = .05$.

a


AFAM
WFWM
AFWF
AMWM
Average boundary
Observer
70
65
60
55
50
45
40
35
30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

b

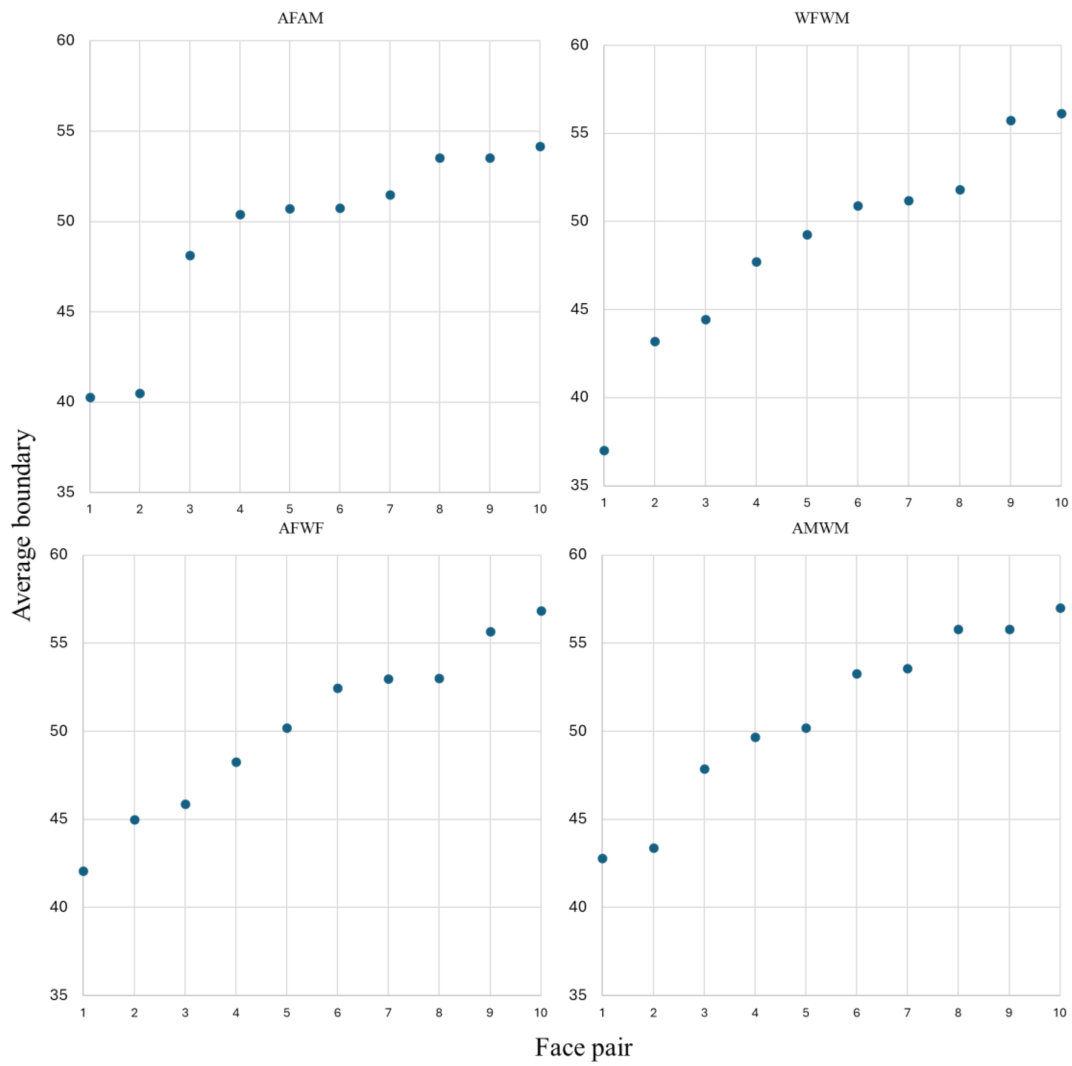


Figure 3.4. Boundary variance from observers and stimuli. a) Average boundary variance across observers for each face set, organized from lowest to highest. The four panels show data for the four face sets: top left is AFAM, top right is WFWM, top left is AFWF, and bottom right is AMWM. b) Average boundary variance across the ten identities within each set, represented in panels the same as a).

3.4.2 Correlations between face pairs and factor analysis

To assess the influence of the individual faces on the category judgments, we examined the correlations for all face pairs (Figure 3.5). Patterns of high correlations suggest that observers were performing in similar ways for a given pair. These high correlations are evident for the race judgments (lower right quadrant of the figure), suggesting that the differences between observers were fairly consistent across the face exemplars; or in other words, that observers were showing "general" differences in how they were categorizing the faces. In contrast, the correlations are weaker for the sex judgments, suggesting that the observer differences for these categories were more dependent on the specific faces being judged.

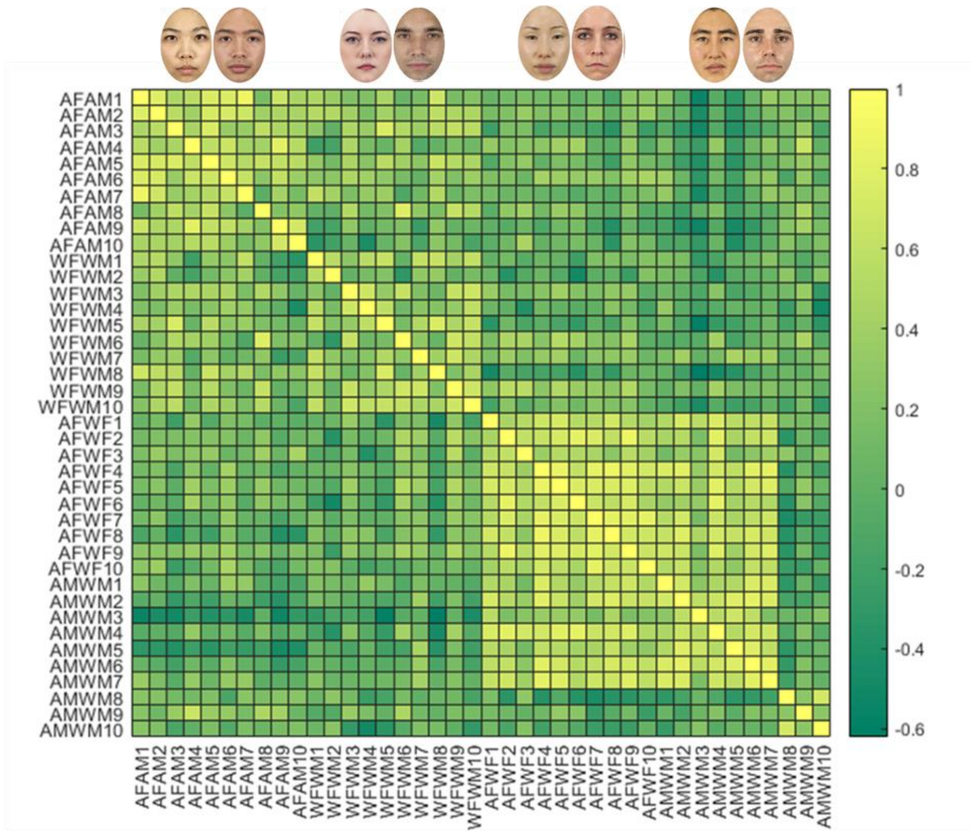


Figure 3.5. Correlation matrix between average responses to fair pairs. Correlations organized in a heat map to represent patterns in response similarity for all face pairs. Large patterns of similarity are seen especially in the categories for race, indicating that response variability is occurring at the level of the categories and not for specific face identity pairs.

To explore these patterns, we conducted a factor analysis of the correlations in observer responses. Factor analysis is a common technique used for dimension reduction by identifying underlying sources of variation in a set of data. A typical analysis involves multiple steps with various options (see Costello & Osborne, 2005; Emery, Volbrecht, Peterzell, & Webster, 2017, 2023). Our factors were computed from the correlation matrix of observer categorization responses and extracted with principal component analysis (PCA). The extracted factors are represented by their factor loadings, which specify the correlation between the factor and each variable. We used the standard

Varimax rotation which favors solutions in which the loadings tend to be very high. Generally, there must be as many factors as variables to represent all the variance, but only some of these account for real variations as opposed to noise. The number of factors to extract was determined by examining their loadings so that the factors with the highest loadings are likely to correspond to meaningful variations. The extracted factors were retained for categorization loadings of 0.6 or higher. This is an approach that has been applied previously in factor analyses (Emery et al., 2017; Peterzell and Teller, 2000; Peterzell, Chang, & Teller, 2000; Webster and MacLeod, 1988).

The first eight factors accounted for 91% of the variance, while the first four factors accounted for 76% of that variance. The loadings for these factors are shown in Figure 3.6, with the face pairs arranged so that the loadings are ordered from lowest to highest. The top panel shows the first four factors. Factor 1 accounts for much of the variation for race judgments for both male and female faces, while factor 2 accounts for the sex judgments for Asian faces. That is, for these factors, the differences between observers were largely consistent across the face pairs. For race, this means each observer was applying a consistent criterion regardless of the sex or identity of the faces (though some face pairs were exceptions). For sex, the factor also suggests that when judging the Asian faces, observers were again applying a consistent criterion independent of the specific identity. In contrast, two factors (factors 3 and 4) emerged for judging the sex of the white faces, implying that observers may have relied on two different strategies or stimulus cues that depended on the identity. Note again that most observers for this experiment were white in race, and all primarily exposed to a diet of white faces, and this dimension was the only one composed of faces (female and male) that were both white.

Including factors 5-8 (Figure 3.6, bottom panel) brings the variance accountability to 91%. These factors capture the variance from outlier faces in the top panel, such as faces 31, 32, and 33, which were not accounted for by Factor 1 but seem to be captured by Factors 7 and 8.

Although four loadings accounted for most of the face set response variations within each category, some face sets fell outside of those factors. Figure 3.7 shows examples of the three face pair exemplars that had low loadings for Factor 1 in the AMWM category compared to face examples that have the highest loadings for that factor. There does not seem to be a conspicuous visual distinction for these face stimuli, so the bases for these differences remain unclear. Figure 3.8 shows examples of the face exemplars with the highest loadings for both Factor 3 and Factor 4, given that they shared loadings within a single face category. Similarly, there do not seem to be visual distinctions driving the separation of factor loadings within this face category.

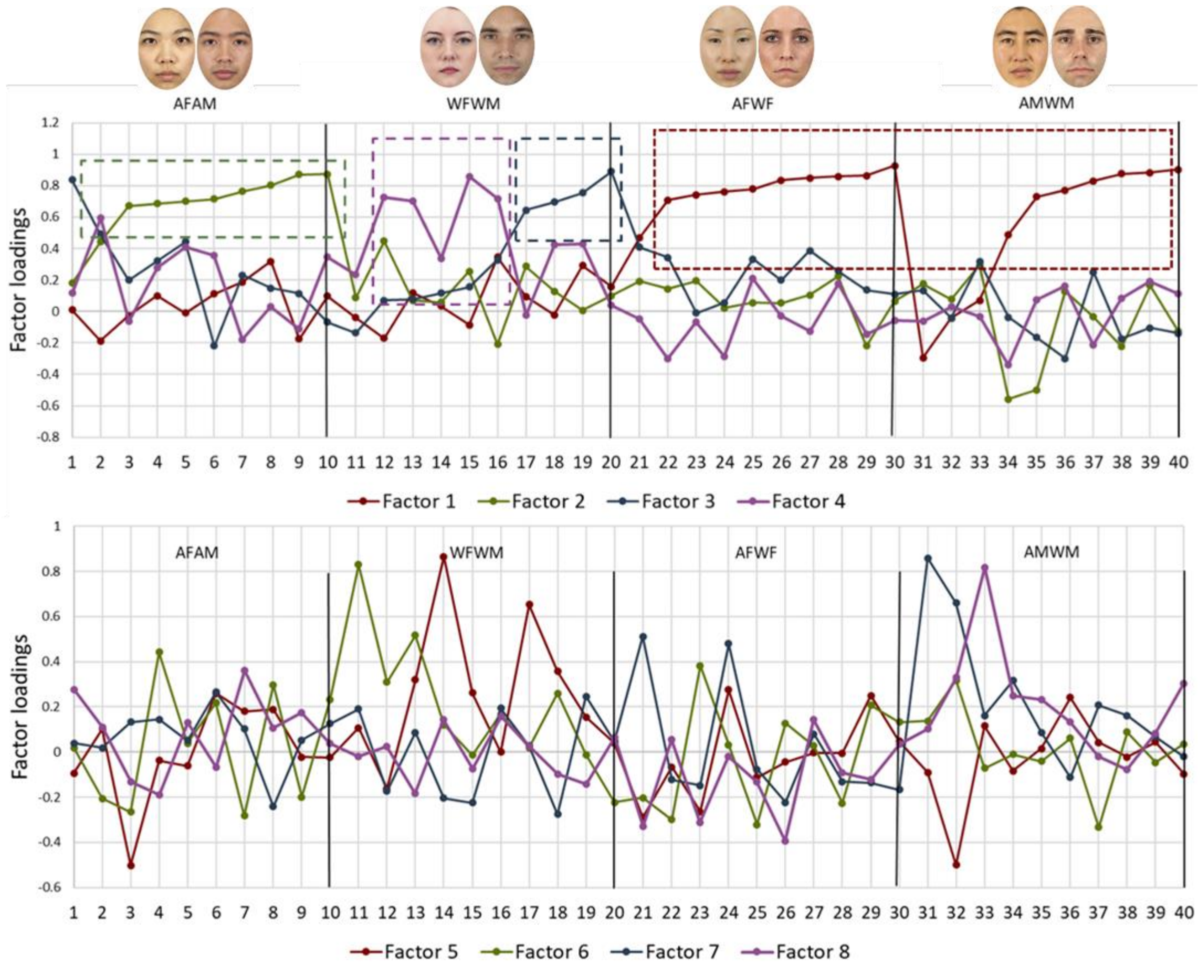


Figure 3.6. Factor loadings by face stimulus pairs. *Top panel* shows factor analysis loadings accounting for 76% of the variability observed in categorization response data. Dashed lines emphasize the four factors within the face categories. Factor one (red) accounted for variability of the race judgments in both sex categories, and factor two (green) accounted for variability in the sex judgements for Asian faces. The sex judgments for Caucasian faces were accounted for by two factors: three (blue) and four (purple). This is interesting as it is the only category with a split in the factor loadings and the only race in-group category for observers. The *bottom panel* shows loadings for faces in the same order as the top panel on the remaining factors, 5-8, which brings the variance accountability to 91%. It is evident these factors do capture the variance from outlier faces in the top panel, such as faces 31, 32, and 33, which were not accounted for by Factor 1 but seem to be captured by Factors 7 and 8.

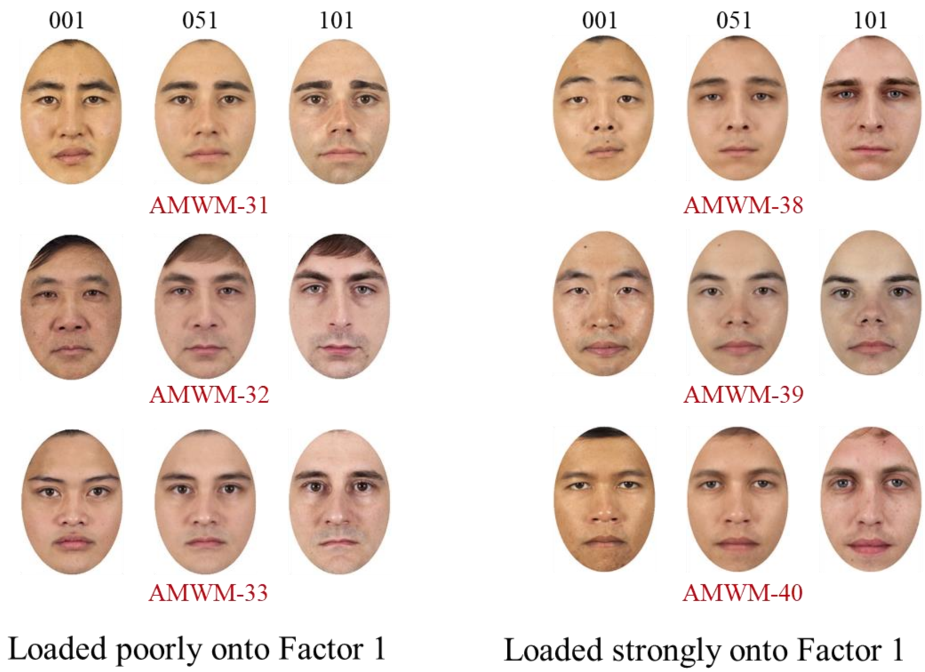


Figure 3.7. Comparison of face pairs with low vs. high loadings for Factor 1. Faces on the *left panel* are stimuli 31, 32, and 33 from Figure 3.6. These three face stimulus sets had the lowest loadings for Factor 1 while the *right panel* shows stimuli 38, 39, and 40, which had the highest loadings for this factor. Each set shows the two original identities (001 and 101) and the 50% morph (051).

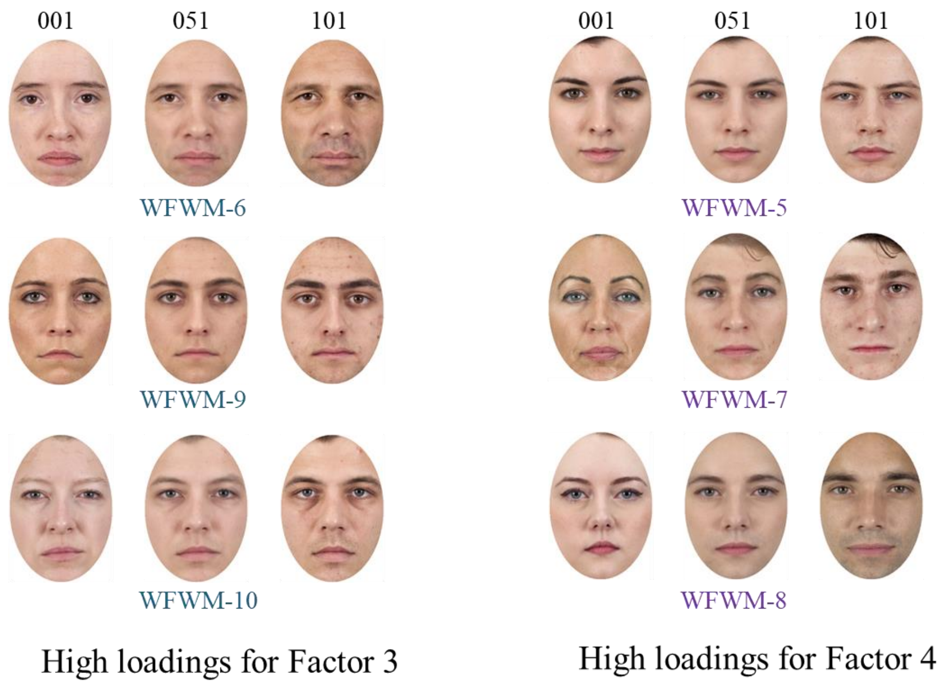


Figure 3.8. Comparison of face pairs with high loadings for Factors 3 and 4. Faces on the *left panel* are stimuli 6, 9, and 10 from Figure 3.6. These three face stimulus sets had the highest loadings for Factor 3. The three sets on the *right panel* are stimuli 5, 7, and 8, which had the highest loadings for Factor 4. Each set shows the two original identities (001 and 101) and the 50% morph (051).

3.5 Discussion

The main aim of the current study was to assess the ways individuals categorize the race and sex of different morphed facial identities. It is apparent that these judgments must in part depend on the properties of the faces being judged, because individual faces can vary widely in how well they represent different facial attributes. However, it is unclear how observers themselves differ, and whether these differences are consistent across faces or instead depend strongly on the specific characteristics of the faces. To assess this, we measured category boundaries for both a large number of observers and real faces.

3.5.1 Correlational patterns of boundary responses

For three of the four face categories examined, we saw strong correlations in observer boundary responses for most of the ten identities. While there were exceptions in the pattern, overall responses were highly correlated across exemplars, indicating that whatever an observer was doing for one face identity set in a category, they seemed to be doing for all sets in that category. This suggests that the differences between observers were consistent across most face pairs, though the mean interobserver biases might vary with the stimulus pair. However, for one of the dimensions (WFWM) observers did not merely differ in a general bias but rather in face-specific biases, similar to a pattern that has been found for trustworthiness where multiple dimensions were pulled from a factor analysis when trying to build a model to account for judgements for a population (Sutherland, Rhodes, Burton, & Young, 2020; Sutherland et al., 2020).

3.5.2 Factors driving variability

A follow up PCA revealed four main factors accounting for 76% of the variance in observer responses. Of importance are potential implications of the breakdown of these factors and what they could reveal about processes operating for categorization in our face groups. Factor one accounted for variability in response to race judgments for both male and female faces. While we do know exactly what that factor is representing, the pattern of loadings suggest that observers were using a consistent strategy for classifying the race that was largely independent of the sex or identity of the face. Factor two emerged for sex judgments, but only for Asian faces. Thus, in this case the judgments were again largely independent of identity but specific to race – i.e. the observers were using different strategies to judge sex in the Asian vs white faces. Finally, for the white race exemplars, there were instead two different factors underlying the judgments, implying that the judgments depended on two different and identity-specific cues. The pattern for this factor is again notable because almost all observers were white in race, and it was also the only dimension restricted to white faces.

Again, we cannot know from PCA alone what is specifically driving responses, but from the breakdown of factors there may be some in-group or other-race effect happening for categorization of the all-Caucasian vs. mixed-race pairings. Note, ingroup and ORE are not synonymous, although are related. Ingroup is much broader than race, and can refer to any group of faces that you relate as being an ingroup to yourself, including judgements on the basis of race, sex, community, club, sport, etc. Race is just one possible component of ingroup and outgroup, and likely would be the most relevant here, although not necessarily the only factor at play. Many perceptual effects have been identified as being associated with ingroup judgements, including stronger holistic processing for faces categorized as being in-group (Hugenberg & Corneille, 2009), larger N170 amplitude responses elicited by ingroup faces (Ratner & Amodio, 2013), a lack of detriment when outer features are removed for judging ingroup vs. outgroup faces (Sporer & Horry, 2011), and maybe most commonly, a perceptual benefit when judging ingroup faces. Further tests of these ideas could include repeating the measurements with an Asian population of observers to see if the factor pattern follows the race of the observer, or to look more generally at group or familiarity effects on the ratings.

Chapter 4

Perceptual vs. response norms in face categorization revealed by individual differences in face adaptation

Kassandra R. Lee, Michael A. Webster

Department of Psychology, University of Nevada, Reno

Acknowledgments

This research has been supported by EY-010834.

Stimuli used were created by collaborators from Chapter 2: Koyo Nakamura[2], Yusuke Nakashima[3], Masami Y. Yamaguchi[3], Katsumi Watanabe[2]

[2] Waseda University, [3]Chuo University

4.1 Abstract

Individuals differ reliably in how they categorize faces for attributes such as ethnicity or biological sex, but the bases for these differences remain poorly understood. We examined whether differences in categorical boundaries for sex reflect differences in criterion or differences in neural coding of faces, by comparing subjective judgments of the category boundary (which we term the “perceptual norm”) with differences in the neutral point for adaptation aftereffects for faces (which we term the “response norm”). Average male and female faces (each formed by blending 10 individual faces) were morphed to form a graded variation from female to male. A staircase procedure was used to identify the category boundary (point at which there is equal probability of classifying the face as female or male), before or after adapting to faces corresponding to different levels along the morph. Prior adaptation to a female (or male) face causes subsequent faces to appear more male (or female). The adapting morph level that does *not* bias the responses therefore reveals the stimulus level that coding of the dimension is normalized or calibrated for. We show that these response norms significantly covary with the individual differences in the perceptual norm (r = 0.7). This suggests that a substantial source of the individual differences in the categorical boundaries for sex is how the neural architecture for representing faces is calibrated, potentially through either innate coding differences or the long-term experience of faces for the observer.

4.2 Introduction

Adaptation is defined as a brief and temporary change in sensitivity or perception when we are exposed to a new stimulus (Webster, 2015). This is a crucial skill of our visual system, and many studies have shown that characterizing the dynamic changes of our visual system in response to different stimuli can provide insight into coding strategies and the ways in which they are calibrated. For example, Webster and Macleod (2011) provide visual examples of how stimuli along one dimension, like faces being judged on sex might be captured by a set of channels tuned in different ways (Partial Figure 4.1 from Webster & Macleod, 2011). These changes are typically characterized by studying aftereffects, or a type of afterimage that arises after the visual system adjusts its sensitivity from viewing a stimulus. There are established hallmark patterns of aftereffect changes, such as their selectivity to stimulus properties and tendency to reflect reduced sensitivity for stimuli similar to an adaptor. While they can be similar, aftereffects also show a range of varying dynamics based on the stimulus to which we are adapted and studying them can reveal something about the response changes in neural mechanisms underlying perception (Webster et al., 2004). Adaptation to low level stimuli, such as pattern orientation yields aftereffects indicative of adaptation at different visual cortical areas (Paradiso, Shimojo, & Nakayama, 1989), while adaptation to color has shown aftereffects selective for different orientations based on the adapted pattern (McCollough Howard & Webster, 2011). Adaptation can be used to explore the processes and dynamics underlying perception of higher-level stimuli, like faces as well.

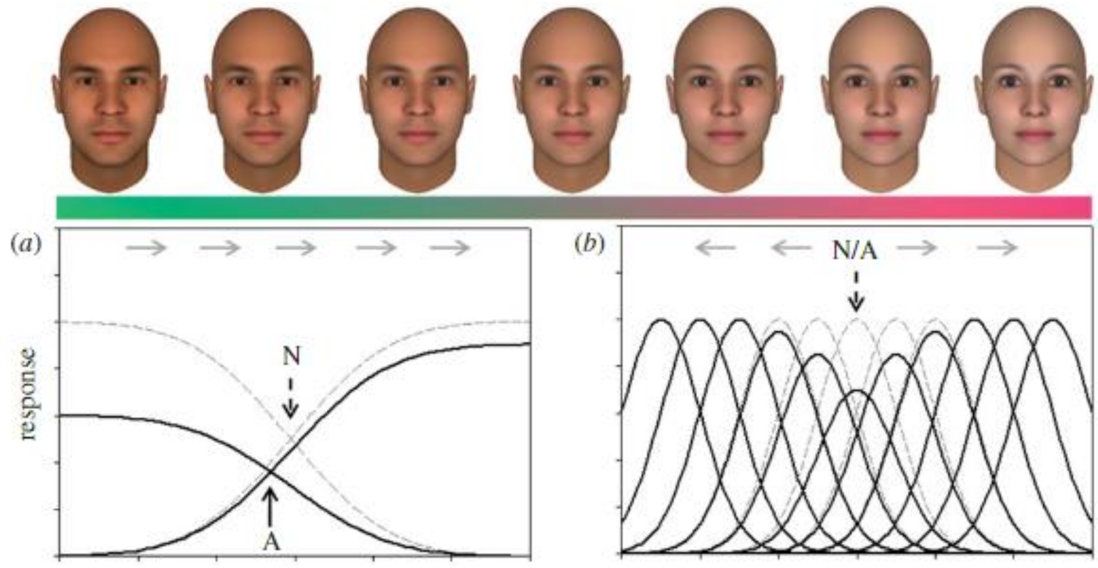


Figure 4.1. Norms and channel coding along a face sex dimension. The sex dimension may be represented by a small number of broadly tuned channels (left) or a larger number of narrowly tuned channels (right). N is a point of neutral adaptation (no shift in the norm) and A represents adapting level. a) The norm is represented by equal responses in two channels. Adapting to a biased stimulus (shifted away from the norm) reduces the response in one of the channels more and shifts the norm away from neutral. This produces a shift in the appearance of all faces in the direction indicated by the arrows. b) Both the stimulus and channels are narrowband. Adaptation reduces the channel response at the adapting level and skews the responses to other stimuli away from the adapting level. There is not a unique norm (both N and A occupy same location). *For details, see Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1571), 1702-1725.*

While aftereffects have been demonstrated at lower perceptual levels for stimuli like orientation or motion, the aftereffects studied for faces are inherently more complex since the adapting stimuli themselves carry multifaceted information about identity, mood, etc. (Leopold, Rhodes, Muller, & Jeffery, 2005). These aftereffects have been shown to possess a robustness, ultimately less affected by changes from the adapting to test stimuli than for lower-level effects. Worth exploring in the future are reasons for this robustness, as it may be related to other interesting components of adaptation, like timescale (Gao, Pieller, Webster, & Jiang, 2022). Leopold et al. (2005) found that face adaptation had a similar trajectory as that to other, lower-level stimuli. Specifically, an aftereffect for judging face identity grew stronger the longer a person was adapted, and the effect dropped off as a function of how long a test face was viewed. This overlap in aftereffect patterns raises questions about the origins of adaptation in the visual stream. While low level stimuli are recruiting early visual areas and more complex stimuli are recruiting later visual areas, the overlap in aftereffect patterns may be evidence that to some degree, adaptation is evoking activation of the entire visual cortex. Additional studies have looked at aftereffects as they are associated with a face memory task, and reported results consistent with the idea that aftereffects can tap into high-level face-space (Dennett, McKone, Edwards, & Susilo, 2012), which is less likely to be corruptible by low-level attributes. Although we understand a lot about the process of adaptation, continuing to investigate it for more complex stimuli will be informative for the ways in which the brain is responding to different stimulus types.

Previous studies have shown figural aftereffects in the perception of faces, whereby people saw test faces as distorted in a direction opposite that of an adapting face. Beyond this, the aftereffects transferred across face stimuli and showed an asymmetry such that adaptation did not occur in the opposite direction, i.e., adapting to an original unedited face did not affect the perception of a distorted face. This provides evidence that adaptation may serve to normalize our perception of faces and is altering sensitivity at a high level of perceptual encoding, beyond low level feature differences (Webster & Maclin, 1999). Additionally, studies have found that adaptation transfers across changes in stimulus size (Zhao & Chubb, 2001) and retinal position (Leopold et al., 2001). Watson and Clifford (2003) also found that aftereffects transfer across changes in stimulus orientation, while Tillman and Webster (2012) showed that aftereffects had strong transfer even across face identity. Yamashita, Hardy, De Valois, and Webster (2005) showed that featural differences are not the basis for adaptation selectivity. Together, these findings provide further support that face adaptation is adjusting sensitivity at higher, face-specific levels of processing, independent of these low-level feature cues.

Another study showed that you can recalibrate preferences for faces by adapting people to distorted faces and observing a subsequent shift in what was perceived as normal and attractive towards that distortion. This result demonstrates that you can rapidly recalibrate preferences by updating the norm via adaptation (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003). We can even adapt to face silhouettes of different genders and view the aftereffect change in front-view faces, revealing common neural mechanisms as the site of adaptation for these stimulus types (Davidenko, Witthoft, & Winawer, 2008). The ubiquitous adapting effects observed for many domains of face perception tell us that adaptation likely serves a critical purpose as it routinely influences face perception in our daily lives. To support this, Webster et al. (2004) studied adaptation effects for natural variations in faces, as opposed to extreme distortions or silhouettes of aforementioned studies. They found that natural stimulus variations, across dimensions like sex, race, and expression, are large enough to cause different states of adaptation in observers. This emphasizes the importance of the diet of faces to which we are all exposed in our daily environments, and may tell us something about the criteria we use to judge faces and the sensitivity of our visual systems to different faces we see.

Of interest is the idea that individual differences in face perception may be linked to individualized norms (Webster, 2015; Webster & Leonard 2008, Sawides et al. 2011, Radhakrishnan et al. 2015). As people move through the world and are exposed to a specific diet of faces, their visual systems generate a norm, established by adaptation. The perceptual experience of individuals will vary relative to this norm, such that individual faces will be represented by how they deviate away from it (Webster & MacLeod, 2011). Norm-based coding provides a good explanation for individual differences in face perception seen across observers, as everyone's norm will be different based on their personalized experience (Rhodes, Robbins, Jaquet, McKone, Jeffery, & Clifford, 2005). If aftereffects can tap into high-level face-space, as reported by Dennett et al., (2012), they may be a useful way to study how individuals code faces. By studying adaptation of the visual system to a norm or stimuli varying in distance from the norm, we can better understand if these norms are shaped by criterion or sensitivity differences of observers. Criteria refers to the way individuals interpret a stimulus, whereas a sensitivity difference speaks to the neural architecture and reflects how our visual systems are calibrated to the environment.

Here, I am interested in investigating the mechanism by which individuals, with established differences in their norm faces, are adapting to faces varying across a dimension of sex. I adapt people to faces varying in their male- or female-ness and see how the aftereffects of their adaptation reflect either a criterion or sensitivity difference.

4.3 Experimental methods

4.3.1 Participants

Participants included 11 observers from Reno, Nevada (5 female, 1 Asian). All participation was with informed consent and followed protocols approved by the University of Nevada, Reno IRB.

4.3.2 Stimuli

Stimuli were morphs of 10 faces (same morphs as in chapter 2) to create averages for a sex judgment category. The faces were frontal-view images of young-adult and white-race of individuals of Swiss nationality (Figure 4.2). The face images were cropped to remove external features.

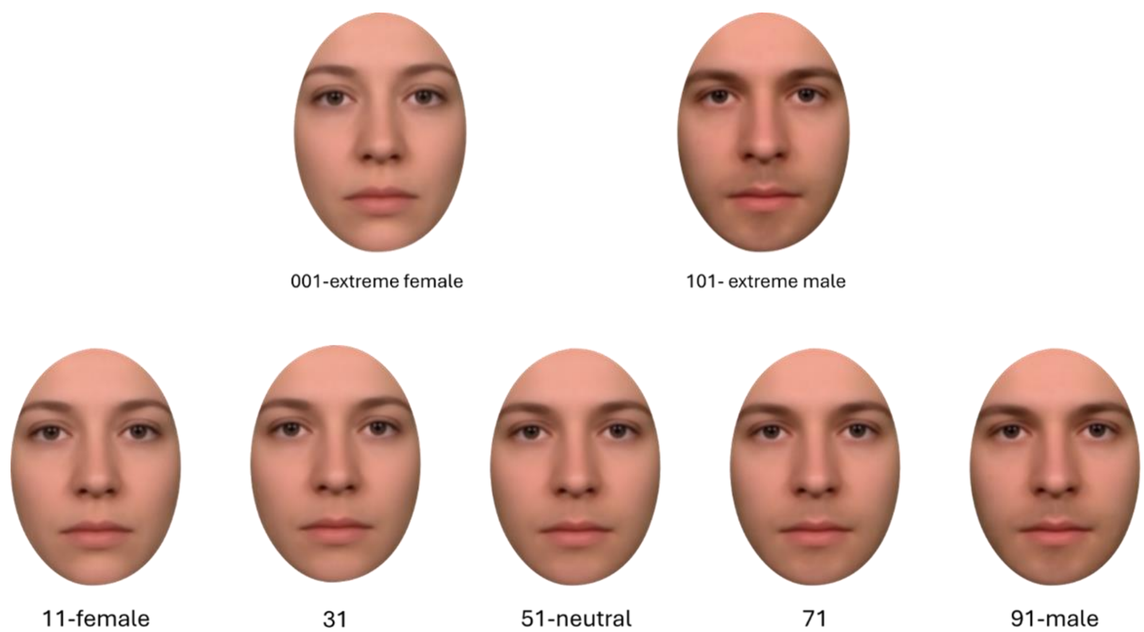


Figure 4.2. Examples of face stimuli for the dimensions of sex. Morph arrays consisted of 100 images spanning the two original faces, fully female and fully male (top), which were assigned a value of 001 or 101. The morph levels for adaptation (11, 31, 51, 71, 91) therefore corresponded to how far along the sequence the image fell between the two original faces.

4.3.3 Procedure

Stimuli were presented in Matlab 2022b on a 32” Cambridge Research Systems Display++ with 1920x1080 resolution and a 120Hz refresh rate. The monitor screen was calibrated with a Photo Research PR 655 spectroradiometer. Observers viewed the display binocularly, seated 80cm away from the monitor, so each individual test face stimulus subtended ~7.5° and each adapt face subtended ~11° of the visual angle.

To estimate the category boundary – at which observers were equally likely to describe the presented face as male or female – we used a one-up one down staircase procedure (Figure 4.3). A morph level was chosen at random to display the initial face for 500 ms. Observers made a forced-choice response to classify the sex. Morph levels for subsequent faces varied in a staircase (i.e. if the response was “male” the next morph shown was incremented in the “female” direction or vice versa), with the category boundary taken as the mean of the final 8 of the total 10 reversals.

In the initial settings, observers judged the faces without a prior adapting stimulus. These settings define their “preadapt” category boundary. Following analogous studies examining individual differences in color perception (Webster & Leonard, 2008), we refer to the “unadapted” settings as the perceptual norm, because it corresponds to their subjective null for the stimulus in the absence of a short-term context and is therefore their intrinsic code for classifying the face.

To examine how this perceptual norm is biased by adaptation, we also measured the category boundaries after adapting to different levels of the morph. As noted, adapting to a male face makes an androgynous face appear more female or vice versa, so there must be some adapting level that does not produce a bias. We refer to this as the response norm because it is the level that maintains the underlying calibration of the neural code (the level at which the adaptation isn't affecting sensitivity). For adaptation, the participant was first shown the adapting face for one minute. Adapting identities ranged from a face that was extremely female to a face that extremely male, for a total of five adapting faces (face 11-very female, faces 31, 51- neutral, 71, 91-very male). Following adaptation, the staircase procedure was repeated and interleaved with 3000 ms top-up of the adapt face to maintain a stable state of adaptation. The adapt stimulus was larger in size than the test faces to prevent biases resulting from adaptation to low-level stimulus features. The first two reversals for all observers were excluded, so final boundary thresholds are calculated from the last eight reversals. All observers received all five levels of adaptation, plus the baseline no adapt condition (always run prior to adapt conditions) six times across two separate days. Results reported are based on the mean of the repeated settings for each observer and condition.

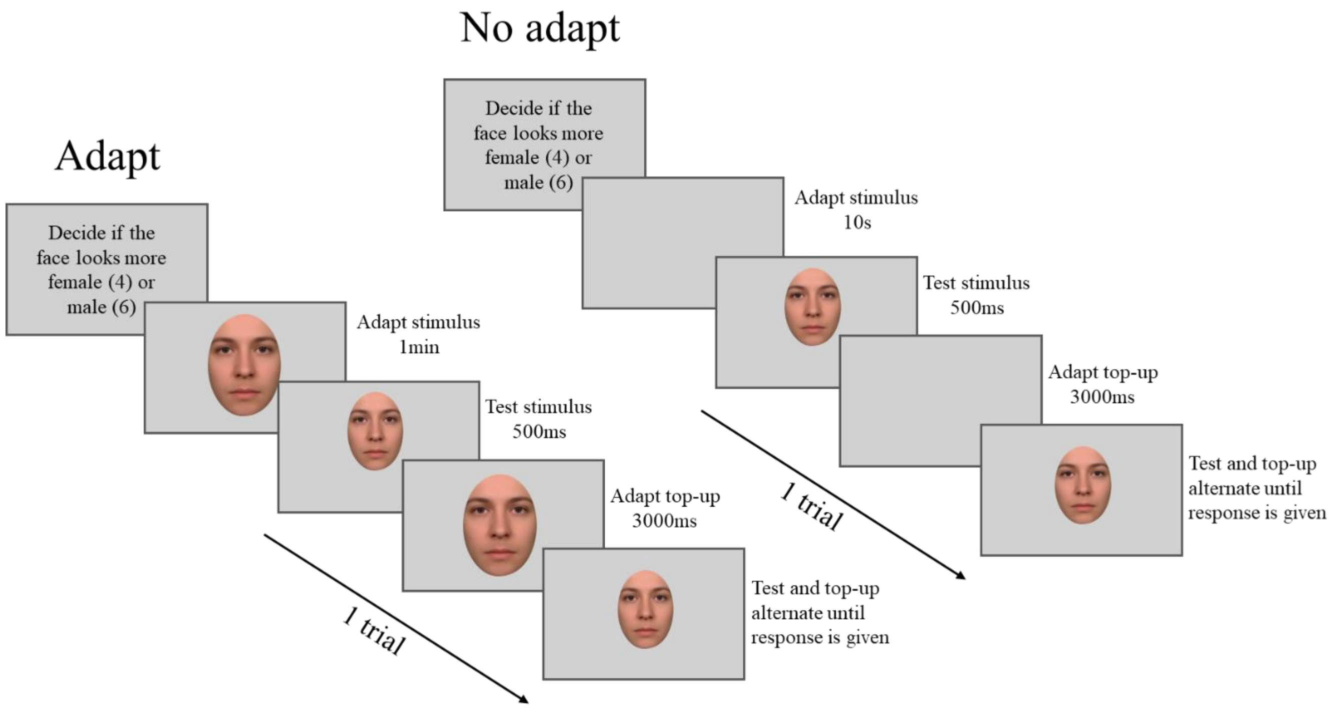


Figure 4.3. Adaptation experimental design. *Left panel* walks through an adaptation trial, starting with instructions then a one minute adapt before alternating a test face and a top up adapt stimulus. Observers respond to the test face, indicating if it looks more male or female. The *right panel* shows a no adapt condition, which was always run prior to an adapt condition. It looks like the adapt condition, except in place of an adapting face, observers view a blank screen.

4.4 Results

4.4.1 Individual differences in category boundary for sex

Comparisons of multiple observer runs showed that between observer variability (M = 8.5, SD = 2.8) was 2.2 times greater than within observer variability (M = 3.8, SD = 1.1), which was significantly different ($t(6) = -3.7$, $p < .01$). Thus, there were reliable inter-observer differences in category boundaries. There was no effect of observer sex, meaning no significant differences in categorization responses for male or female observers. This contrasts with the observer differences reported by Webster et al. (2004) for a larger sample of observers.

4.4.2 Adaptation and shifts in the category boundary

Figure 4.4 plots the settings before and after adaptation for each of the observers tested. For all observers the category boundary for identifying a face as male or female varied with the adapt level. Specifically, as the adapt level became more male, their category boundary also shifted to a more male level of the morph (implying that the original boundary appeared more female) or vice versa. For most observers the adapted settings span the setting chosen before adaptation. However, for one observer (010) the post-adapt values were far outside their pre-adapt boundary and strongly shifted toward the female exemplar relative to other subjects. We are not sure of the basis for these settings but excluded this observer from further analysis because of the large disparity between their pre- and post-adapt judgments.

For the remaining observers we ran a linear regression analysis on each individual's settings. This showed that there were significant adaptation effects of all observers (all $p <= .02$). The linear fits provided a good approximation to the settings (all $R^2 >= 0.84$). We therefore used this fit to estimate the pattern of aftereffects for each observer, and specifically, to estimate the adaptation level that did not produce a change in individual pre-adapt category boundaries. This corresponded to the intersection of the adapting trend line with the pre-adapt setting (as indicated by the horizontal black line for each observer). Importantly, with the exception of observer 7, the response norms and perceptual norms within each subject were similar, as shown by comparing the red triangles with the cross point in figure 4.4.

To formally evaluate this correspondence, we correlated the observers' (pre-adapt) perceptual norms with their (post-adapt) response norms (Figure 4.5). This revealed a moderately strong correlation (r = 0.7, p =.003). A linear fit line to the values had a slope of 1.3, suggesting that observer response norms varied more than their perceived norms. However, much of this difference is driven by the results for one observer (S7), for which as noted there was not a close correspondence between their pre- and post-adapt neutral points. With this observer excluded the slope changed to 1.1 and the correlation increased (r = 0.8, p = .003).

Settings across observers can become more similar to each other as they normalize to a common stimulus. We tested the variance for observers in their no adapt condition vs. variance across the five post-adapt settings and found that post-adapt variance was 1.7x lower compared to no adapt variance. This supports the idea that observers were normalizing and becoming more similar in response to the five adapt face levels. Face level 11 was an outlier here, as when variance was assessed at that level alone it yielded variance higher than the no adapt setting.

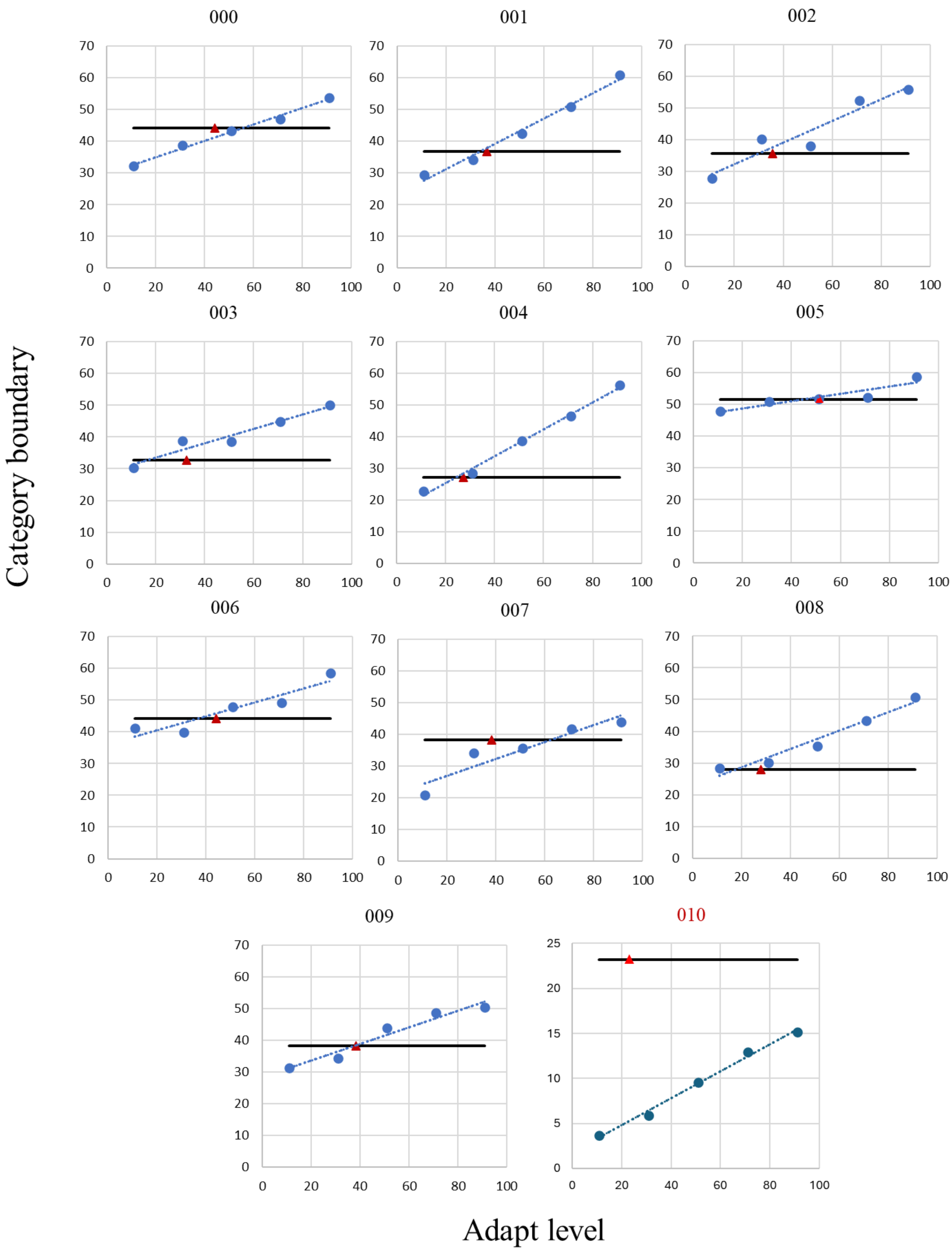


Figure 4.4 Individual observer threshold responses for all adapt levels. Observers denoted at the top of each graph: 000-010. Observer 10 was excluded from further analyses as their post adapt

level never crossed their pre-adapt. Blue dots represent the threshold responses across all adapt levels, which trends similarly for all observers and has a slope significantly greater than 0, indicating adaptation. Dashed blue line represents a linear fit to each observer's data. All R^2 values were .84 or above. The horizontal black line is each observer's no adapt null value. The cross point of the no adapt with the data fit indicates each observer's response norm – i.e. the adapt level that does not produce a change in the pre-adapt or perceptual null. Concordance of the perceptual and response null predicts the lines should intersect at the level of the perceptual null, as indicated by the red triangles.

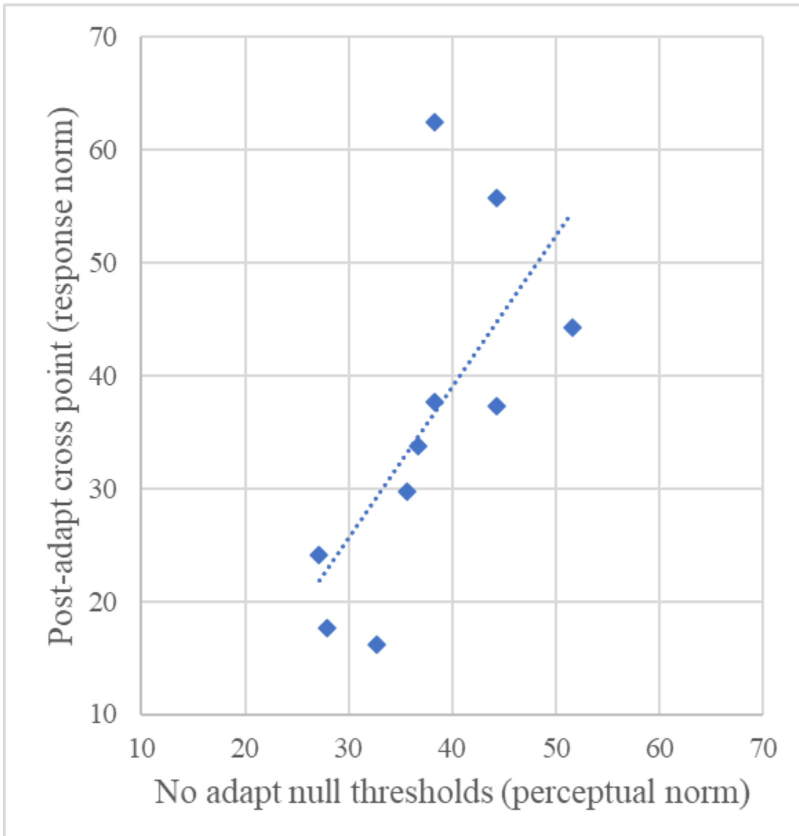


Figure 4.5. Correlation of individual perceptual norms and response norms. Blue diamonds show threshold data and dashed line is linear fit. A strong correlation between observer perceptual (x-axis) and response (y-axis) norms show that they covary.

4.5 Discussion

In summary, the aim of the current study was to see if individual differences in categorizing the sex of face images reflect differences in criteria or in the underlying neural representation of faces. We adapted observers to different levels of face morphs and measured subsequent aftereffects based on their responses to the morphed stimuli. As expected, observers' category boundary was shifted by the adaptor face sex (Webster et al., 2004), except for some intermediate adapt level that did not produce a shift. The correlation between this neutral adapt level – or response norm – with their pre-adapt category boundary – or perceptual norm – suggests that at least part of the differences in how observers categorized the faces reflects differences in the neural responses to the faces, and not merely different criteria applied to a common neural representation.

4.5.1 Negative aftereffects

All observers showed significant adaptation to the varying levels of the adaptor face, consistent with the robust aftereffects demonstrated in previous studies (Webster et al., 2004; Webster & MacLeod, 2011; Gao et al., 2022). For example, when adapted to the original female face, observers saw the original boundary level as too male and required a physically more female face in order to perceive the test face as neutral. This specific type of post-adaptation bias has been demonstrated for many aspects of faces, including identity (Leopold et al., 2001; Leopold et al., 2005), race (Little, DeBruine, Jones, Waitt, 2008; Webster et al., 2004), sex (Little, DeBruine, & Jones, 2005), and attractiveness (Rhodes et al., 2003) (see review by Webster and MacLeod, 2011).

While the current study morphed faces to create a male-female sex dimension, other studies have shown that you can also adapt these categories independently, causing sex-contingent aftereffects for male and female faces (Little, deBruin, & Jones, 2005). For example, viewing faces transformed along dimensions of identity or masculinity increased preferences for novel faces, but only when the sex of the adapt and test faces were congruent. These findings may provide insight into the neural codes underlying the perception of male and female faces, indicating distinct populations of neurons encoding them. This provides interesting implications for the current study as the differences in where observers draw their sex boundaries may indicate how variable these populations of neural codes are for individuals and tell us that these populations of neurons are susceptible to tuning on a short timescale.

4.5.2 Sensitivity differences drive individual variability

A strong correlation between observers' no adapt null thresholds and their post-adapt cross points suggest that sensitivity changes are driving at least part of their individual boundaries for face sex, as opposed to the percepts being driven by criterion differences. Adaptation provides an avenue to distinguish criterion vs. sensitivity changes by measuring the neutral adapt level of the face that does not produce an aftereffect. This stimulus level is congruous with individual response norms at the neural locus of sensitivity change, as the lack of aftereffect indicates that level does not change responses within response channels (Webster, 2011). This approach was first used to assess how color vision is normalized. Webster and Leonard (2008) found that the response norms for chromatic adaptation were close to the subjective achromatic point for individual observers, and that this correspondence held in both the fovea and near-periphery, even though spectral sensitivity at the two locations differed because of differences in macular screening pigment. This suggested that the two locations were adapted or normalized for the same physical stimulus even though the retinal stimulus was very different. Moreover, this normalization for color coding must occur at or prior to the site of the adaptation, which for color is largely at the site of the cone receptors. Subsequent studies applied the same logic to examine whether the visual system is adapted to the level of ambient blur in the retinal image (Sawides, de Gracia, Dorronsoro, Webster, & Marcos, 2011; Radhakrishnan, Dorronsoro, Sawides, Webster, & Marcos, 2015). These studies showed that the neutral level for blur adaptation also corresponded to the level of subjective image focus, and moreover that both the subjective focus and adaptation was based on mechanisms showing complete binocularity, implicating a cortical site for the effects.

To our knowledge, the present study is the first to apply this paradigm to assess high level perceptual judgments and aftereffects. The current results are similar in showing a correspondence between perceptual and response norms. This suggests that individual differences in the perception of the sex of faces partly depends on differences in the underlying, long-term calibration or adaptation state, potentially driven by individual differences in exposure to a particular diet of faces. By assessing variance in boundary thresholds before and after adaptation, we found that individuals had less variability overall after adaptation, indicating they were normalizing to a common stimulus.

It is still of interest to determine where adaptation transitions from being brief sensitivity changes to more complex shifts in functional mechanisms or cellular responses that are indicative of neural plasticity or perceptual learning (Webster, 2011; Lu, Yu, Watanabe, Sagi, & Levi, 2009). Functionally defining adaptation becomes increasingly difficult as we find that our visual systems adapt at abstract levels, demonstrating a large variety of changes based on individual experience. For example, Dils and Boroditsky (2010) found that people can adapt to motion extracted from mental imagery derived from linguistic descriptions. Other studies have found that you can adapt to sex from viewing biological motion (Jordan, Fallah, & Stoner, 2006), adapt to faces from viewing a body (Ghuman, McDaniel, & Martin, 2010), and adapt to faces from only imagining them (D'Ascenzo, Tommasi, & Laeng, 2014). A better understanding of aftereffects, their time course, and the ways in which they cause changes in sensitivity will be necessary to understand the adaptability of our visual systems and exploit that as a tool to unveil the ways in which we encode perceptual features across all complexity levels.

In this regard, it is important to consider more closely the underlying nature of the representation of faces and how this is altered by adaptation. Many forms of adaptation reflect a renormalization for the adapting stimulus, so that this stimulus itself appears more neutral or closer to the norm. For example, adaptation to a color causes that color to appear less saturated, or more like the gray norm. For faces, this would correspond to a male (or female) face becoming more androgynous the longer we adapt to it. Many aspects of face coding are consistent with this account (Webster and MacLeod, 2011). However, for gender, Storrs and Arnold (2012) have argued that the adaptation to the facial attribute of gender is instead a "contrastive" effect, where the adapt face itself remains similar but other faces appear more different to it. By this account, each adapting face should have looked the same after adapting, and this could lead to an alternative explanation where the response norm corresponded to the perceptual norm simply because the adaptor at that level would remain the same in appearance. However, there are a number of arguments against this account. First, these kinds of contrastive effects should occur when the bandwidths or tuning of the individual channels are narrow relative to the coding dimension (Webster, 2011). This in turn predicts that aftereffects should be strongest for faces near the sex boundary while weaker for faces farther from the boundary. Instead, we found a monotonic and largely linear increase in the magnitude of the aftereffect with increasing adapt-boundary distance, and this pattern is considered a signature of norm-based coding (Webster and MacLeod, 2011; Valentine, 1991; Loffler, Yourganov, Wilkinson, & Wilson, 2005; Leopold, Bondar, & Giese, 2006). Moreover, while some drop-off occurs at larger separations (Zhao, Series, Hancock, & Bednar, 2011), these are outside the range of natural variation in faces, and thus unlikely to reflect response changes in the mechanisms directly involved in face coding (Pond et al., 2013; Robbins, McKone, & Edwards, 2007). These different effects have been debated with regard to whether the coding is norm-based (e.g. based on the relative activity of two opposing channels) vs. multichannel (e.g. based on the peak response among a population of channels) (Webster and MacLeod, 2011; Blakemore & Campbell, 1969; Calder, Jenkins, Cassel, & Clifford, 2008; Jenkins, Beaver, & Calder, 2006). However, in practice these models do necessarily lead to different patterns of aftereffects. For example, chromatic adaptation is “multichannel” in the sense that it varies the relative activity of the three class of cones. However, because the cones have broad spectral sensitivities relative to the visible light spectrum, the adaptation leads to renormalization of color perception rather than a contrast effect at the adapting wavelength. Similarly, the monotonic increases in adapt strength are consistent with a renormalization of the perceived magnitude of biological sex in the face images, even if that magnitude is coded by a distribution of channels rather than an explicit opponent code. And the fact that this renormalization is again relative to observer’s neutral point is suggestive of the idea that the differences in how sex is judged reflects differences in how faces are actually *seen*, and not simply how they are interpreted.

Chapter 5

General discussion, limitations, and future directions

General discussion

Current studies

Across three experimental studies I unveiled individual differences in the ways that people perceive and categorize faces. Some differences are reflective of population-level variability, though interestingly within-region differences were most prominent. These differences were stable and persisted across different face identities within and across their race and sex categories. Judgements also seem to be related to individual neutral points or norms, and following adaptation, aftereffects reflected a sensitivity change of observers based on that norm, indicating perception of the sex of faces partly depends on differences in the underlying, long-term calibration or adaptation state, potentially driven by individual differences in exposure to a particular diet of faces.

Cultural differences in face perception have been studied through the lens of the other race effect (ORE) (Feingold, 1914). This work is useful for identifying population-level differences that drive perceptual processes. For example, Masuda et al. (2008) found that Japanese individuals incorporate context into their emotional judgments of faces, while Westerners do not. This work yielded behavioral as well as eye-tracking differences between the groups and shed light on the way social aspects of culture can drive differences. Additional work identified many interacting factors driving differences in judging female faces, including the ethnicity and sex of both subject and assessor (Voegeli et al., 2021a, b). Population level differences yield important information about societally driven variation, although many stable differences in percepts within a culture exist as well. The current study found that differences in face categorization were larger within a region than across. This finding is consistent with one that exists in the field of color vision research, in which color terms were more consistent across cultures than within (Berlin & Kay, 1969). It may be the case that categories such as race and sex are not unlike color term categories. Across cultures, these categories are stable, like the regularities in the coding of color, but within a culture there are a multitude of factors driving differences in the ways we are sensitive to category boundaries.

As an attempt at understanding the factors that influence differences in face perception, my second study expanded the categorization judgments to different identities for the race and sex categories. Individual differences persisted, with variability occurring at the level of the category and not for specific pairs of faces. The differences between observers were consistent across most face pairs, though the mean interobserver biases varied with the stimulus pair. This means that variability is not dependent on the individual identity, but instead persists for a range of faces. This generalizability is worth noting, as it can be informative of a degree of commonality of underlying mechanisms driving perception across different facial identities (McCaffery et al., 2018). Further, studies have shown that there are differences in the variability of judgments observed when people are assessing familiar versus unfamiliar faces, so more detailed exploration of face identity is needed (Megreya & Bindemann, 2013; Burton, White, & McNeill, 2010; Woodhead & Baddeley, 1981).

In a final experiment, I utilized an adaptation paradigm to unveil sensitivity differences across observer category boundaries. More investigation is needed to determine if the pattern of aftereffects we observed could be due to contrastive effects of faces other than the adapt, but currently the pattern of our results suggests a renormalization of the adapting stimulus. This provides evidence that people are at least partially adapting to faces based on their own sensitivity differences surrounding norms, likely updated from their long-term adaptation to the world around them.

Adaptation paradigms and namely face aftereffects have been useful at providing insight to the ways that individuals reference a face-space when making judgments about face stimuli. Leopold et al. (2001) showed that adaptation shifted perception along a trajectory that increased selectivity of a test face and impaired recognition of other faces, indicating that face encoding utilizes the norm in face space to facilitate perception. Robbins, McKone, and Edwards (2007) tested several experimental conditions that support a norm-based face coding model by showing that aftereffects shifted in the direction of a norm face as opposed to just away from the adapter. Webster and Maclin (1999) also found figural aftereffects from viewing distorted faces and reported that adaptation may be important for understanding configural properties of face perception, especially since we are sensitive to such properties of faces.

Because face-space is a good culprit of variation in face perception across observers, continued studies on the aftereffects of faces are important to gain a better understanding of the mechanisms driving the ways that we see faces and how those may differ across individuals. Dennett et al. (2012) identified important criteria for the aftereffects that can provide the most information on face-space, including studying aftereffects specifically related to facial coding as opposed to lower-level configural properties. This is especially important because many studies have identified effects of adaptation on low level features (McCollough Howard & Webster 2011; Foster 2011; Paradiso et al. 1989; Mather et al. 2008), so it will remain important to parse those out from adaptation to face-level features for future studies.

Improving the study of individual differences

Individual differences are a useful tool to elucidate broader perceptual concepts. Wilmer (2007) emphasized the ways in which the study of individual differences has allowed researchers to test cognitive theories (Underwood, 1975). For example, Wilmer (2008) describes the importance of considering individual differences in what he referred to as Nature's experiments, or experiments that encompass a full range of natural variation in an individual's ability, as it is affected by unique genes and environment. While the sources of individual differences are still being explored for many perceptual tasks, for color vision, the study of individual differences has been promising as insight into the functional organization and genetic underpinnings to which Wilmer refers. He makes the case that perception beyond the retina should also be explored through an individual difference lens. Wilmer (2008) lays out five central principles to making use of individual differences in vision science. I won't go into extensive detail, but will mention three of these briefly, including experiments that have exemplified these principles.

The first principle is considering data from both natural and lab experiments (Cronbach, 1957) to determine potential differences being driven by each. Kosslyn et al. (2002) developed such a study on individual differences that utilized variation often ruled out as noise. By combining thorough laboratory results with historic research on theories and biological mechanisms, they reiterated the point that multiple approaches as well as group and individual data should be combined to yield the most informative outcomes.

A second principle involves identifying and accounting for noise in data. For a long time, individual differences have been overlooked as noise, probably because non-systematic, uninformative noise does occur in data, and we cannot draw conclusions until this noise is reduced. This can arise from many things, including observer biases, instrument error, even observer mood. These are confounding and prevent the study of true individual differences, which in turn are systematic and intrinsic to the individual (Mollon et al., 2017). Methods such as attenuation correction help to compare correlated measures and rule out measurement error as the source of variation (Schmidt & Hunter, 1996).

The third principle I'll mention is addressing questions of relation between mechanisms. Individual differences are powerful in this domain, as they harness overlapping and distinct manipulations on the visual system via difference-based correlations. Underwood (1975) studied individual differences to uncover relations between independent variables of interest and utilized this to support causal theories.

These principles are worth consideration when studying individual differences. Because the differences available to uncover are vast, so are the ever-evolving methods and analyses for best capturing them.

Limitations/future directions

The amount of variability that can be identified across individuals is seemingly infinite. To classify that variability and try to connect it to underlying mechanisms is a feat that we've only begun to undertake in the field of vision science. For that reason, our attempts are not without limitations. Papers like Wilmer (2008), Bosten (2022), and Mollon et al. (2017) should be heavily referenced when building individual difference experiments. For example, the current study could benefit from consideration of more real-world, natural components. That could mean using more naturalistic stimuli, adapting to a larger variety of faces, or testing a more diverse population of observers. The current findings are a good start towards understanding individual differences in face categorization, and the goal moving forward should always be to improve methods aimed at capturing individual differences.

The current studies utilize faces, which are one of the most multidimensional and complex stimuli we perceive. Common issues surrounding face perception studies include deciding what type of face stimuli to use. The faces here were morphed identities, and in some cases even morphed across many more than two identities. Studies have shown that averaged faces are viewed as more attractive (Valentine, Darling, & Donnelly, 2004; Langlois & Roggman, 1990). Although not an attractiveness study, it is worth noting that averaged faces may be perceived differently than individual identities. Similarly, there is a tradeoff between real and generated faces. For all experiments here, the morphed identities were composed of real face images from databases. While using real identities provides more real-world validity and generalizability, the downside is that you don't control as many featural aspects of the stimuli. Some of our faces had makeup, while others had wrinkles, varying hairlines, etc. These are all features that may affect the way people make categorical judgments about faces and should be considered when selecting real vs. generated faces, although new research is still investigating if humans can distinguish AI-generated faces from real ones (Shen, Richard, Webster, O'Toole, Bowyer, & Scheirer, 2021; Bray, Johnson, & Kleinberg, 2023). The study of face categories itself is becoming increasingly complex. Many aspects of race and sex (or gender) can be viewed on a spectrum (e.g., mixed race, nonbinary identification), and the ways that individuals think about these spectra varies widely and is likely changing based on societal exposure. Future research should aim to collect more qualitative measures from observers to gain an understanding of the ways they think about these topics and are exposed to them in their everyday lives, and well as make strides to create experiments that can capture these differences (Kozan, 2020; Fisher et al., 2020).

Only part of the current studies was run in Tokyo, Japan. The comparison of Reno and Tokyo observers was fruitful and would go a long way in providing additional clarification to the Reno data collected in studies two and three. More data should be collected in Tokyo to shed light on some of the PCA potential in/outgroup effects from study two, specifically. Additional follow ups should be done on the adaptation experiment to parse out if the aftereffects are truly reflective of sensitivity differences, like our data suggest, or if there may be some contrastive effect causing our response norm to coincide with the perceptual norm.

Significance/application

While no experiment will perfectly capture all there is to know about variability in face processing, the current research makes positive strides towards identifying variability for face categorization and attempting to understand the underlying mechanisms driving that variability. Maybe the most important outcome from studies like these is thinking about why this research matters and where it can be applied.

Given that we know a lot about face processing from behavioral measures, many recent studies have turned to neuroimaging techniques to investigate mechanisms in the brain. This is true for studies of individual differences as well. Furl, Garrido, Dolan, Drier, and Duchaine (2011) associated findings of individual differences in face processing ability to core face processing regions in the brain. Seghier and Price (2018) considered variance in brain functioning to be informative data representative of plasticity as opposed to simple noise. This approach had ramifications for characterizing typical vs. atypical brain functioning, revealing cognitive strategies that underpin tasks, and predicting recovery after brain damage. As we progress behavioral studies to uncover mechanisms driving variability, neuroimaging studies can begin to use this to inform studies aimed at identifying neural underpinnings of individual differences. The combination of behavioral and neuroimaging studies will go farther in providing the full story of individual differences and identifying their usefulness.

A major area of growing research in vision science is artificial intelligence, or AI. Whether it’s being used to provide guidance to humans for face processing (Crum, Boyd, Bowyer, & Czajka, 2023; Boyd, Tinsley, Bowyer, & Czajka, 2023), or to generate synthetic media (Whittaker, Kietzmann, Kietzmann, & Dabirian, 2020), AI is becoming more ubiquitous as a tool in science. While it is a powerful tool, AI is trained based on the things we know about the human visual system (Boyd et al., 2023). That means our studies of how humans perceive faces are being used to inform AI, and our understanding of individual variation will feed into this as well. Advances in computer vision based on knowledge of how humans process faces have already been utilized by leveraging caricatures to train algorithms (Davis and Hand, 2022; Davis, Lingenfelter, McElhinney, Sengupta, & Hand, 2023). This research is informed by the knowledge that caricatures can be identified more efficiently (Sun, Wang, & Tang, 2014). Additionally, given what we know about the ORE, it is important that we make strides towards building variability/diversity into AI models (Cavazos, Noyes, & O'Toole, 2019; O'Toole, & Castillo, 2021; O'Toole, Deffenbacher, Abdi, & Bartlett, 1991; Scheuerman, Paul, & Brubaker, 2019). This means ensuring that training of these algorithms is driven by a wide variety of face races, ages, sex, etc. so that biases in these categories do not skew performance.

Final conclusions

There is still much to understand about perception at the level of the individual. When we take the time to look for patterns in individual differences, we uncover a wealth of knowledge about underlying cognitive and sensory processes of human perception. These differences exist for many types of stimuli in our perceptual environment. As we investigate them for higher level stimuli, like faces, identifying the basis for differences becomes increasingly complex and informative. The key to capitalizing on these differences is appropriate measurement and analysis to parse true, systematic differences from noise. While no one study meets all the requirements for best practices and principles, many, including the current studies, have begun to yield compelling and invaluable insights into the variability in human perception and experience.

References

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the other-race effect. *Current directions in psychological science, 22*(3), 173-178.

Asano, Y., Fairchild, M. D., & Blondé, L. (2016). Individual colorimetric observer model. *PloS one*, *11*(2), e0145671.

Balsdon, T., Summersby, S., Kemp, R. I., & White, D. (2018). Improving face identification with specialist teams. *Cognitive Research: Principles and Implications*, *3*, 1-13.

Berlin B, Kay P. 1969. Basic Color Terms: Their Universality and Evolution. Berkeley: Univ. Calif. Press.

Blakemore, C. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol. 203*, 237-260

Bliss, C. I. (1934). The method of probits. *Science, 79*(2037), 38-39.

Bosten, J. M. (2022). Do you see what I see? Diversity in human color perception. *Annual review of vision science, 8*, 101-133.

Boyd, A., Tinsley, P., Bowyer, K., & Czajka, A. (2023, June). The value of ai guidance in human examination of synthetically-generated faces. *In Proceedings of the AAAI Conference on Artificial Intelligence, 37*, No. 5, pp. 5930-5938.

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect ‘deepfake’images of human faces. *Journal of Cybersecurity, 9*(1), tyad011.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. Journal of Personality and *Social Psychology, 42*, 673– 681. http://dx.doi.org/10.1037/0022-3514.42.4.673

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications*, *3*, 1-3.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior research methods, 42*(1), 286-291.

Calder, A. J., Jenkins, R., Cassel, A. & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *J. Exp.Psychol. Gen. 137*, 244–261. (doi:10.1037/0096-3445.137.2.244)

Cavazos, J. G., Noyes, E., & O'Toole, A. J. (2019). Learning context and the other-race effect: Strategies for improving face recognition. *Vision research, 157*, 169-183.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation, 10(1), 7.

Cramer, J. S. (2002). The origins of logistic regression.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist, 12*(11), 671.

Crum, C. R., Boyd, A., Bowyer, K., & Czajka, A. (2023). Teaching ai to teach: Leveraging limited human salience data into unlimited saliency-based training. arXiv preprint arXiv:2306.05527.

D'Ascenzo, S., Tommasi, L., & Laeng, B. (2014). Imagining sex and adapting to it: Different aftereffects after perceiving versus imagining faces. *Vision Research, 96*, 45-52.

Davidenko, N., Witthoft, N., & Winawer, J. (2008). Gender aftereffects in face silhouettes reveal face-specific mechanisms. *Visual Cognition*, *16*(1), 99-103.

Davis, S. R., & Hand, E. M. (2022). Improving face recognition using artistic interpretations of prominent features: Leveraging caricatures in modern surveillance systems. In Intelligent Video Surveillance-New Perspectives. IntechOpen.

Davis, S. R., Lingenfelter, B., McElhinney, K., Sengupta, S., & Hand, E. M. (2023, September). CarVer: Setting the Standard for Face Verification with Caricatures. In 2023 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-10). IEEE.

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological science*, *23*(11), 1279-1287.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences, 107*(37), 16396-16400.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576-585.

de-Wit, L., & Wagemans, J. (2014). Individual differences in local and global perceptual organization.

Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017). Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision research, 141*, 51-65.

Feingold, G. (1914). The Influence of Environment on Identification of Persons and Things. of Crim*. Law and Criminal, 5*, 39-51.

Filkowski, M. M., Olsen, R. M., Duda, B., Wanger, T. J., & Sabatinelli, D. (2017). Sex differences in emotional perception: Meta analysis of divergent activation. *Neuroimage, 147*, 925-933.

Fisher, A. D., Ristori, J., Castellini, G., Cocchetti, C., Cassioli, E., Orsolini, S., ... & Gavazzi, G. (2020). Neural correlates of gender face perception in transgender people. *Journal of Clinical Medicine, 9*(6), 1731.

Foster DH. 2011. Color constancy. *Vis. Res. 51*:674–700.

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of cognitive neuroscience, 23*(7), 1723-1740.

Gao, Y., Pieller, J., Webster, M. A., & Jiang, F. (2022). Temporal dynamics of face adaptation. *Journal of Vision, 22*(11), 14-14.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review*, *19*, 847-857.

Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Current Biology, 20*(1), 32-36.

Gignac, G. E., Shankaralingam, M., Walker, K., & Kilpatrick, P. (2016). Short-term memory for faces relates to general intelligence moderately. *Intelligence*, *57*, 96-104.

Hammond BR, Caruso-Avery M. 2000. Macular pigment optical density in a Southwestern sample. Investig. Ophthalmol. *Vis. Sci. 41*(6):1492–97.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition, 106*, 1017–1027. http://dx.doi.org/10.1016/j.cognition.2007.04.002

Held, R. (1989). Perception and its neuronal mechanisms. *Cognition, 33*(1-2), 139-154.

Howard, C. M., & Webster, M. A. (2011). McCollough effect. *Scholarpedia*, *6*(2), 8175.

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science, 33*(6), 1173-1181.

Jenkins, R., Beaver, J. D. & Calder, A. J. (2006). I thought you were looking at me: direction-specific aftereffects in gaze perception. *Psychol. Sci. 17*, 506–513. (doi:10. 1111/j.1467-9280.2006.01736.x)

Jordan, H., Fallah, M., & Stoner, G. R. (2006). Adaptation of gender derived from biological motion. *Nature neuroscience, 9*(6), 738-739.

Kaiser, P. K. (1988). Sensation luminance: a new name to distinguish CIE luminance from luminance dependent on an individual's spectral sensitivity. *Vision research, 28*(3), 455-456.

König, A., & Dieterici, C. (1886). The modern development of Thomas Young's theory of colour-vision. *Report of the British Association for the Advancement of Science, 56*, 431-439.

Kosslyn, S. M., Cacioppo, J. T., Davidson, R. J., Hugdahl, K., Lovallo, W. R., Spiegel, D., & Rose, R. (2002). Bridging psychology and biology: the analysis of individuals in groups. *American psychologist, 57*(5), 341.

Kozan, E. (2020). The effect of gender roles on gender identity, sexual orientation and subjective perception of attractiveness (Doctoral dissertation, YAŞAR UNIVERSITY).

Langlois, J., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*, 115-121.

Lee, K. R., Richardson, A. J., Walowit, E., Crognale, M. A., & Webster, M. A. (2020). Predicting color matches from luminance matches. *JOSA A, 37*(4), A35-A43.

Lee, K.R. & Webster, M.A. Environmental influences on color vision. In Encyclopedia of Color Science and Technology (Second Edition) (2020). R. Luo, Ed. (Springer).

Leopold, D. A., Bondar, I. V. & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature 442*, 572–575. (doi:10.1038/nature04951)

Leopold, D. A., Rhodes, G., Müller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 897-904.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*, 89–94.

Little A. C., DeBruine L. M., Jones B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. Proceedings of the Royal Society B: Biological Sciences, 272, 2283–2287.

Little A. C., DeBruine L. M., Jones B. C., Waitt C. (2008). Category contingent aftereffects for faces of different races, ages and species. *Cognition, 106*, 1537–1547.

Loffler, G., Yourganov, G., Wilkinson, F. & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nat. Neurosci. 8*, 1386–1390. (doi:10.1038/nn1538)

Lu, Z. L., Yu, C., Watanabe, T., Sagi, D., & Levi, D. (2009). Perceptual learning: Functions, mechanisms, and applications. *Vision Research, 50*, 365–367.

MacLeod, D.I.A. and Webster, M.A. (1983). Factors influencing the color matches of normal observers. In Mollon, J.D. and Sharpe, L.T. (Eds.), Colour Vision: Physiology and Psychophysics, Academic Press, London.

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

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of personality and social psychology, 94*(3), 365.

Mather G, Pavan A, Campana G, Casco C. 2008. The motion aftereffect reloaded. *Trends Cogn. Sci. 12*:481–87.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive research: principles and implications*, *3*, 1-15.

McCollough Howard C, Webster MA. 2011. McCollough effect. *Scholarpedia 6*:28175

McGugin, R. W., Van Gulick, A. E., & Gauthier, I. (2016). Cortical thickness in fusiform face area predicts face and object recognition performance. *Journal of cognitive neuroscience*, *28*(2), 282-294.

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology, 25*(1), 30-37.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3.

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research, 141*, 4-15.

Namba, S., Sato, W., Nakamura, K., & Watanabe, K. (2022). Computational process of sharing emotion: An authentic information perspective. *Frontiers in Psychology, 13*, 849499.

O'Toole, A. J., & Castillo, C. D. (2021). Face recognition by humans and machines: three fundamental advances from deep learning. Annual Review of Vision Science, 7, 543-570.

O'Toole, A. J., Deffenbacher, K., Abdi, H., & Bartlett, J. C. (1991). Simulating the 'other-race effect'as a problem in perceptual learning. *Connection Science, 3*(2), 163-178.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*, 208-224.

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An 'other-race effect'for categorizing faces by sex. *Perception, 25*(6), 669-676.

Paradiso, M. A., Shimojo, S., & Nakayama, K. (1989). Subjective contours, tilt aftereffects, and visual cortical organization. *Vision research*, *29*(9), 1205-1213.

Peterzell, D. H., & Kennedy, J. F. (2016). Discovering sensory processes using individual differences: A review and factor analytic manifesto. *Electronic Imaging, 28*, 1-11.

Peterzell, D. H., Chang, S. K., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red–green and luminance-modulated gratings: psychophysical data from human infants. *Vision Research, 40*(4), 431-444.

Peterzell, D. H., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red–green and luminance-modulated gratings: psychophysical data from human adults. *Vision Research, 40*(4), 417-430.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, *115*(24), 6171-6176.

Pokorny J, Smith VC, Lutze M. 1987. Aging of the human lens. *Appl. Opt. 26*(8):1437–40.

Pond, S., Kloth, N., McKone, E., Jeffery, L., Irons, J., & Rhodes, G. (2013). Aftereffects support opponent coding of face gender. *Journal of Vision, 13*(14), 16-16.

Radhakrishnan, A., Dorronsoro, C., Sawides, L., Webster, M. A., & Marcos, S. (2015). A cyclopean neural mechanism compensating for optical differences between the eyes. *Current Biology*, *25*(5), R188-R189.

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.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological science*, *14*(6), 558-566.

Rhodes G., Jeffery L., Watson T. L., Jaquet E., Winkler C., Clifford C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14*, 2119–2123.

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C. W. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In *Fitting the mind to the world: Adaptation and after-effects in high-level vision*. Oxford University Press.

Robertson, D. J., Mungall, A., Watson, D. G., Wade, K. A., Nightingale, S. J., & Butler, S. (2018). Detecting morphed passport photos: A training and individual differences approach. *Cognitive research: principles and implications*, *3*, 1-11.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PloS one, 11*(2), e0150036.

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: adapter position effects and neural models. Journal of Experimental Psychology: *Human Perception and Performance, 33*(3), 570.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review, 16*(2), 252-257.

Sawides, L., de Gracia, P., Dorronsoro, C., Webster, M. A., & Marcos, S. (2011). Vision is adapted to the natural level of blur present in the retinal image. *PloS one*, *6*(11), e27031.

Scheuerman, M. K., Paul, J. M., & Brubaker, J. R. (2019). How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-33.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological methods, 1*(2), 199.

Seghier, M. L., & Price, C. J. (2018). Interpreting and utilising intersubject variability in brain function. *Trends in cognitive sciences, 22*(6), 517-530.

Shen, B., RichardWebster, B., O'Toole, A., Bowyer, K., & Scheirer, W. J. (2021, December). A study of the human perception of synthetic faces. In 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021) (pp. 1-8). IEEE.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision research, 49*(22), 2705-2739.

Sporer, S. L., & Horry, R. (2011). Recognizing faces from ethnic in-groups and out-groups: Importance of outer face features and effects of retention interval. *Applied Cognitive Psychology, 25*(3), 424-431.

Storrs, K. R., & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision research, 64*, 7-16.

Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1898).

Sutherland, C. A., Burton, N. S., Wilmer, J. B., Blokland, G. A., Germine, L., Palermo, R., ... & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences, 117*(19), 10218-10224.

Sutherland, C. A., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality?. *British Journal of Psychology, 111*(2), 215-232.

Tillman, M. A., & Webster, M. A. (2012). Selectivity of face distortion aftereffects for differences in expression or gender. *Frontiers in Psychology, 3*, 14.

Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & Cognition, 42*, 863– 875. http://dx.doi.org/10.3758/s13421-014-0409-y

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30*(2), 128.

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.

Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review, 11*, 482-487.

Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V., Shackelford, T. K., & Fink, B. (2021a). Cross-cultural perception of female facial appearance: A multi-ethnic and multi-centre study. *Plos one, 16*(1), e0245998.

Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V., Shackelford, T. K., & Fink, B. (2021b). Differences between perceived age and chronological age in women: A multi-ethnic and multi-centre study. *International Journal of Cosmetic Science, 43*(5), 547-560.

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*(1), 102.

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition, 144*, 91–115. http://dx.doi.org/10.1016/j.cognition.2015.07.011

Watson, T. L., & Clifford, C. W. G. (2003). Pulling faces: An investigation of the face-distortion aftereffect. *Perception, 32*, 1109–1116.

Webster, M. A. (2015). Visual adaptation. *Annual review of vision science*, *1*, 547-567.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557-561.

Webster, M. A., & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Josa a*, *25*(11), 2817-2825.

Webster, M. A., & MacLeod, D. I. (1988). Factors underlying individual differences in the color matches of normal observers. *JOSA A, 5*(10), 1722-1735.

Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1702-1725.

Webster, M. A., & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic bulletin & review*, *6*(4), 647-653.

Webster, M. A., Mizokami, Y., & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network: Computation in neural systems*, *18*(3), 213-233.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision research*, *37*(23), 3283-3298.

Werner, J. S., Peterzell, D. H., & Scheetz, A. J. (1990). Light, vision, and aging. *Optometry and Vision Science, 67*(3), 214-229.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS one*, *9*(8), e103510.

Whittaker, L., Kietzmann, T. C., Kietzmann, J., & Dabirian, A. (2020). "All around me are synthetic faces": the mad world of AI-generated media. *IT Professional, 22*(5), 90-99.

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial vision, 21*(6).

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26*(3), 225-230.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. *Cognitive neuropsychology*, *29*(5-6), 360-392.

Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, Nakayama K, Duchaine B. Human face recognition ability is specific and highly heritable. Proceedings of the National Academy of Sciences of the United States of America. 107: 5238-41. PMID 20176944 DOI: 10.1073/Pnas.0913053107 (b)

Wilmer J, Germine L, Chabris C, Chatterjee G, Williams M, Nakayama K, Duchaine B. (2010). A genetic basis for face memory: evidence from twins *Journal of Vision. 10*: 563-563. DOI: 10.1167/10.7.563 (a)

Wilmer, J. B., Germine, L. T., & Nakayama, K. (2014). Face recognition: a model specific ability. *Frontiers in Human Neuroscience*, *8*, 769.

Wilmer JB, Germine L, Williams MA, Nakayama K, Chabris CF, Duchaine BC. (2009). Genetic and environmental contributions to memory for faces: A twin study *Journal of Vision. 9*: 509509. DOI: 10.1167/9.8.509

Wilmer, J. B., & Nakayama, K. (2007). Two distinct visual motion mechanisms for smooth pursuit: evidence from individual differences. *Neuron, 54*(6), 987-1000.

Wilmer J, Russell R, Bronstad M, Kwok H, Anthony S, Germine L. Disagreements about the attractiveness of faces arise largely from past experiences: evidence from twins. *Journal of Vision. 13*: 854-854. DOI: 10.1167/13.9.854

Woodhead, M. M., & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition, 9*, 368-370.

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics*, *22*(3), 333-343.

Yamashita J. A. Hardy J. L. De Valois K. K. Webster M. A. (2005). Stimulus selectivity of figural aftereffects for faces. Journal of Experimental Psychology: *Human Perception and Performance, 31*, 420–437.

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in human neuroscience*, *8*, 562.

Zhao, L., & Chubb, C. F. (2001). The size-tuning of the face-distortion aftereffect. *Vision Research, 41*, 2979–2994.

Zhao, C., Seriès, P., Hancock, P. J., & Bednar, J. A. (2011). Similar neural adaptation mechanisms underlying face gender and tilt aftereffects. *Vision research, 51*(18), 2021-2030