Title

Development of a Holistic Processing Face Recognition Training

Author

Simon-Jennings, Deja Nicole

Publication Date

2024

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Development of a Holistic Processing Face Recognition Training

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Deja Nicole Simon-Jennings

June 2024

Dissertation Committee:
- Dr. Jimmy Calanchini, Chairperson
- Dr. Steven Clark
- Dr. Annie Ditta

The Dissertation of Deja Nicole Simon-Jennings is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would first like to thank my committee members, Jimmy, Steve, and Annie, for their time and expertise. It has sincerely been a pleasure to learn from you all!

I would also like to individually thank my advisor, Jimmy. Thank you for encouraging me to explore what “tickles my brain”, and for providing the training that has enabled me to do so. I am grateful for your investment in my growth and success, and would not be the scholar I am today without your mentorship.

Finally, I would like to thank the Graduate Division for granting me the Dissertation Completion Fellowship, which has allowed me to focus on my dissertation work this final quarter.

Dedication

I dedicate this dissertation to my family – thank you all for your endless support and encouragement throughout my academic journey. Jazz, thank you for keeping me grounded during this process. It hasn’t been easy, and I’m grateful for your patience and encouragement through it all. Dad, thank you for always believing in me and reminding me to enjoy the journey. I truly wouldn’t be here without you. Dori, thank you for being one of my biggest cheerleaders and for all our Disney days together. I love and appreciate you all so much!

ABSTRACT OF THE DISSERTATION

Development of a Holistic Processing Face Recognition Training

by

Deja Nicole Simon-Jennings

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, June 2024
Dr. Jimmy Calanchini, Chairperson

Face recognition errors occur frequently, with consequences that range from personal embarrassment to eyewitness misidentification. Established interventions have taken a variety of approaches in attempts to improve face recognition, yet they have lacked in their capacity for practical use. With this in mind, I created an application-oriented training program in an effort to improve face recognition in real-world contexts. Specifically, I designed the training to teach individuals how to process faces holistically, or in terms of how facial features spatially relate to one another. After developing the training, I conducted three experiments to assess the general efficacy of the training, examine its capacity to improve recognition over time, and compare its impact against two established other-race face recognition interventions.

Experiment samples consisted of 196 to 320 participants, whom I recruited through the UC Riverside Psychology subject pool and CloudResearch Connect. In all experiments, participants completed a baseline recognition memory task, followed by the training or an alternate condition (matched control in Experiments 1 and 2, individuation or cross-race effect awareness in Experiment 3), then completed another recognition memory task. In Experiment 2, participants completed two additional recognition memory tasks 24 hours and 1 week after the manipulation. Memory strength, operationalized as accuracy on the recognition memory task, was compared before versus after the manipulation to determine whether the training produced an improvement in recognition memory ability.

Across the experiments, multilevel modeling revealed that the training did not lead to improved face recognition ability. Instead, training participants generally displayed poorer recognition memory ability after the manipulation compared to their average recognition ability at baseline. Slight fluctuations in recognition ability had returned to baseline levels after one week (Experiment 2), and only the previously established interventions – not the training – led to improved other-race face recognition (Experiment 3). The training may have incited depletion and fatigue among participants, which I seek to address in future research. In future work I will also measure the extent to which the training promotes holistic processing.

Table of Contents

List of Tables

Experiment 1

Experiment 2

Experiment 3

List of Figures

Experiment 1

Experiment 2

Experiment 3

Introduction

Humans possess a unique expertise for recognizing faces (Diamond & Carey, 1986); indeed, the cognitive mechanisms involved in face recognition develop early in childhood, unlike many other cognitive abilities that develop gradually into adulthood (McKone et al., 2012). And yet, we still experience errors in face recognition all the time, from everyday interpersonal blunders (McKone et al., 2023) to more serious mistakes like eyewitness misidentification (Innocence Project, 2020). Given the ramifications of such errors, society would undoubtedly benefit if people could learn how to better recognize faces.

Existing interventions to improve general face recognition have targeted populations such as children with autism (Kouo & Egel, 2016; Tanaka et al., 2010) and individuals with prosopagnosia (Bate et al., 2015; DeGutis et al., 2014), while other lab-based interventions have targeted specific face-recognition deficits within the general adult population (e.g., for outgroup faces: Tanaka & Pierce, 2009; Malpass et al., 1973). Several interventions for the general adult population have also attempted to make participants' face recognition more effective by boosting *holistic processing* (Dolzycka et al., 2014; Hussain et al., 2009; Malpass, 1981; Robbins & McKone, 2003; Sporer, 1991; Woodhead et al., 1979). Also referred to as "configural processing" or "structural encoding", holistic processing involves encoding a face as a whole and capturing unique subtleties in the shape and spatial relationships between features (Maurer et al., 2002; McKone & Yovel, 2009). Holistic processing is implicated in accurate face recognition (Tanaka & Farah, 1993), so to the extent that a given intervention could improve holistic processing, it would be expected that individuals would demonstrate improved face recognition as a result. However, to my knowledge, no existing holistic processing interventions have been application-oriented, or designed for transference to real-world contexts.

The goal of the present research was to develop and test an educational training designed to improve face recognition in everyday life, specifically by promoting holistic processing. With three studies, I aimed to 1) develop a holistic processing training and test its capacity to improve face recognition, 2) examine the effects of the training on face recognition over time, and 3) compare the efficacy of the training against other interventions designed to improve other-race face recognition.

Theories of General Face Cognition

Face cognition refers to the cognitive systems and processes involved in perceiving and recognizing faces (Chernorizov, 2016). According to the most comprehensive model of face cognition, face processing begins with perceiving a face, or extracting structural "codes" that characterize the face's features and their configuration (i.e., structural encoding, also known as holistic processing; Bruce & Young, 1986). Structural codes are stored within a *face recognition unit*, or a mental representation of that face. Whether the perceiver recognizes a previously-seen face will depend upon the degree of overlap between its existing face recognition unit and the structural codes captured during perception of the face.

An expansion upon this framework proposed a three-factor model of face cognition, highlighting the distinction between face perception, face memory (another term for face recognition), and speed of face cognition (Wilhelm et al., 2010). Importantly, this model illustrates that individuals naturally differ in their ability to perceive the configural information of a face, recognize faces they have learned, and execute these processes expediently. Studies comparing monozygotic and dizygotic twins offer convincing evidence for the heritability of face cognition; across multiple measurements of face perception and recognition, the correlation between monozygotic twins' performance was consistently more than double the correlation between dizygotic twins' performance (Wilmer et al., 2010; Zhu et al., 2010). However, Zhu and colleagues (2010) reported that heritability only accounted for approximately 25-39% of the variability in face recognition, indicating that environmental factors play a substantial role in face recognition as well.

Theories of Other-Race Face Cognition

Over the past several decades, cognitive psychologists have sought to understand the mechanisms and brain regions involved in general face recognition. Meanwhile, social psychologists have spent several decades investigating how face recognition can be influenced by race, focusing on a robust phenomenon called the *cross-race effect*. The cross-race effect is characterized by increased difficulty recognizing other-race faces relative to same-race faces, adding a layer of nuance to general face recognition errors. For example, eyewitness misidentification is a leading cause of innocent people's imprisonment, especially when the witness is of a different race than the accused; Black men are most susceptible to such false accusation and conviction (Innocence Project, 2020).

Literature on the cross-race effect has pointed to two potential mechanisms believed to drive the difference in recognition for racial ingroups versus outgroups. One mechanism is based on perceptual processes, assuming that people better recognize same-race faces due to extensive experience in perceiving racial ingroup members (Michel et al., 2006). The other mechanism is rooted in motivational processes, assuming that people better recognize same-race faces due to increased motivation to attend to racial ingroup members (Levin, 2000; Sporer, 2001). In more recent years, new theoretical explanations have proposed a joint contribution of these mechanisms (Hugenberg et al., 2010) and postulated that cultural norms (e.g., the societal status of a racial group) may impact recognition as well (Hinzman et al., 2022; Simon et al., 2023; Wan et al., 2015). Collectively, these perspectives offer some examples of the role that the environment may play in face recognition: who we are surrounded by, who is personally relevant to us, and the cultural norms in our society are all environmental factors that may influence our ability to perceive and recognize faces.

The Importance of Holistic Processing

Frameworks of general and race-specific face cognition both suggest that accurate face recognition hinges on the employment of holistic processing during perception (Bruce & Young, 1986; Tanaka & Farah, 1993). There has been debate regarding what precisely is entailed in holistic processing - however, I and others argue that faces are comprised of parts, such that holistic processing involves encoding features themselves (e.g., their shape), as well as their relationship to other facial features (Maurer et al., 2002; McKone & Yovel, 2009). When features are encoded holistically, they are mentally represented in relation to their surroundings rather than as isolated entities (Diamond & Carey, 1986; Maurer et al., 2002); it is this configural information that enables a perceiver to distinguish one face from another.

For example, perhaps you perceive that your colleague, Bob, has a thick mustache. Encoding this feature in isolation is not diagnostic (i.e., informative for recognizing Bob), because many individuals have thick mustaches. If you perceive that below Bob's mouth takes up a significant proportion of his lower face below his thick mustache, your holistic mental representation of this feature – how it relates to its surroundings – is more diagnostic, thus useful for recognizing Bob's unique face.

With regard to race, participants display greater holistic processing for same-race than other-race faces (Rhodes et al., 1989; Tanaka et al., 2004), suggesting that we possess greater perceptual attunement for faces of our ingroups. Indeed, perception of unfamiliar faces is characterized by equal encoding of both internal (e.g., nose width) and external (e.g., hair color) cues, whereas perception of familiar faces involves encoding of internal cues more exclusively (Bruce & Young, 1986). To the extent that a perceiver is familiar with perceiving same-race faces, they will be more inclined to encode internal cues that convey configural information about the face – storing this more diagnostic information would subsequently enhance their likelihood of recognizing that face at a later point.

In sum, though a small handful of studies have failed to show that holistic processing is implicated in face recognition (Konar et al., 2009; Verhallen et al., 2017), most work reports evidence of their relationship.

Interventions Targeting General Face Recognition

Interventions aimed at improving general face recognition have various distinctions, from their target audience (e.g., children, adults, individuals with and without face processing deficits) to the amount of time they require (e.g., one hour to several months). With regard to their content, many interventions have focused on altering how participants perceive faces; how a face is encoded directly impacts how well it can be later recognized (Bruce & Young, 1986), so it is intuitive that interventions would target this aspect of face cognition.

Some interventions have attempted to teach holistic processing by training participants to perceive inverted faces (Hussain et al., 2009; Robbins & McKone, 2003), however these attempts were unsuccessful in improving subsequent recognition of faces. Other approaches have sought to promote more fruitful encoding by asking participants to make judgments of faces (a task that incites deeper processing; Sporer, 1991) or presenting faces from multiple viewpoints to strengthen perceivers' mental representation of them (Dolzycka et al., 2014), but both of these trainings resulted in minimal change.

Lastly, other interventions have targeted participants' awareness of variability in facial features (Malpass, 1981; Woodhead et al., 1979). Adapted from Penry (1971), these interventions involved teaching participants about variability in different feature characteristics. For example, training material about eye variability highlighted differences in color, eyelid type (e.g., hooded, lidless, etc.), and size. In addition to learning about an assortment of feature variations, participants practiced identifying which variations different faces possessed. Surprisingly, participants demonstrated worse recognition relative to baseline (Malpass, 1981) and control participants (Woodhead et al., 1979) in these studies.

Though most interventions seeking to improve face recognition in normal adult populations have been unsuccessful, the same cannot be said about interventions developed for specialized populations. For example, children with autism have adopted analytic and holistic processing strategies to better recognize faces (Tanaka et al., 2010), and a young adult with prosopagnosia experienced improved face perception as a result of an online intervention (Bate et al., 2015). Most notably, DeGutis et al. (2014) demonstrated in a longitudinal case study that a woman with prosopagnosia learned to effectively extract configural information from faces. After developing the ability to engage in holistic processing over the course of several months, the woman was ultimately able to recognize faces at a level on par with control participants.

Interventions Targeting Other-Race Recognition

Like interventions designed for specialized populations, trainings intended to improve other-race face recognition have also had more success. One approach used in several interventions involves enhancing participants' ability to individuate other-race faces (Elliott et al., 1973; Goldstein & Chance, 1985; Tanaka & Pierce, 2009). In these experiments, participants learned to associate individual other-race faces with unique letters or digits, over the course of one or multiple training sessions. Findings regarding the efficacy of this approach have been mixed; some studies reported successful improvement of other-race face recognition (Elliott et al., 1973), even several months after the training had concluded (Goldstein & Chance, 1985). In contrast, others reported a mix of reliable and descriptive improvements (Tanaka & Pierce, 2009).

Another intervention involved informing participants about the existence of the cross-race effect prior to a face recognition task (Hugenberg et al., 2007). After learning about the phenomenon, participants received the following instructions: “Do your best to try to pay close attention to what differentiates one particular face from another face of the same race, especially when that face is not of the same-race as you… Remember, pay very close attention to the faces, especially when they are of a different race than you in order to try to avoid this Cross Race Effect” (Hugenberg et al., 2007, p. 337). The authors posit that these instructions motivated participants to individuate other-race faces. Hugenberg and colleagues (2007) have demonstrated success with this technique, as have others (Rhodes et al., 2009); however, not all replication attempts have been successful (Cruz et al., 2023; Wan et al., 2015).

Though multiple interventions have led to improved other-race face recognition immediately after, Malpass (1982) points out that these were often short-term improvements that generally did not last over time (Lavakras et al., 1976; Malpass et al., 1973). Collectively, the literature on face recognition interventions is quite mixed. Most attempts to improve general face recognition among normal adults have been unsuccessful; but on the other hand, trainings designed to address various types of face-related deficiencies have had more success.

The Current Research

The limited number of trainings aimed at improving holistic processing have not been ideal for various reasons, with the most critical drawback being their lack of practical use. For example, though performance on face inversion paradigms has been used as a metric of holistic processing, inverted faces do not exist in the real world; therefore, interventions centered around training people to recognize inverted faces arguably do not equip participants with skills they can readily transfer to practical scenarios. Additionally, though the holistic training in DeGutis et al. (2014) improved face recognition for a woman with prosopagnosia, the extensive time entailed in the intervention would not be practical for most laypeople.

For my dissertation, I aimed to create an online training program in which participants could learn to encode faces holistically. In doing so, my hope was that individuals would be able to apply the concepts learned to their daily lives, and ultimately become better at recognizing faces. My training is most similar to the Penly (1971) approach implemented in Woodhead et al. (1979) and Malpass (1981), but with some key distinctions. Whereas their interventions focused on features in isolation, I sought to facilitate the encoding of features in a holistic way; specifically, I attempted to draw participants' attention to how variability within features impacts the spatial relationships (i.e., distance, position) with other elements of the face. I also focused my training on features with substantial variability across ethnic and racial groups (Fang et al., 2011), so that the skills learned could be applicable to a wide range of faces. Additionally, to promote engagement and deeper learning, I designed the training to include an assortment of activities and reflective prompts.

For recognition errors in applied contexts to be reduced, there is a need for a face recognition training that imparts practical strategies individuals can implement in their daily lives. After developing a holistic training program that I felt accomplished this goal, I conducted three experiments that assessed its efficacy in varying ways: by comparing to a matched control, measuring recognition over time, and comparing to other established face recognition interventions.

Experiment 1

I conducted Experiment 1 to test the efficacy of my training. Participants began by completing a recognition memory task to assess their baseline memory. Next, they completed their randomly-assigned experimental condition (i.e., holistic training or matched control). Finally, they completed the recognition memory task once more with new faces. I predicted that training participants would display a more pronounced improvement in recognition memory from pre-manipulation to post-manipulation than matched control participants who did not complete the training.

The primary goal of this intervention is to improve recognition memory for faces of all races, but previous research suggests that the intervention effects would be moderated by target face race. I anticipated that participants in both conditions would display better recognition memory for White than Black faces prior to the manipulation. Participants in this study were undergraduates at the University of California, Riverside, where the population consists largely of Asian and Latinx students. Though all stimuli in this experiment would be racial outgroups to most participants, I have found in my own previous research that American participants consistently better recognize White than Black faces, even among participants (i.e., Asian, Latinx) for whom both White and Black faces are racial outgroups (Simon et al., 2023).

After the manipulation, participants in both conditions could display improved recognition memory for Black faces – due to the holistic processing skills learned and practiced among those in the training condition, or due to the increased attention to specific facial features among those in the control condition. Crucially, I expected that this post-manipulation improvement in recognition memory for Black faces would be more pronounced among training, relative to control, participants, due to previous evidence showing that holistic processing corresponds with greater recognition accuracy than featural processing (Rhodes et al., 2006).

For White faces, I expected post-manipulation recognition memory to mirror my prediction for Black faces. Specifically, though I expected both conditions to display improved recognition memory for White faces, I predicted that there will be a more pronounced improvement among training participants compared to control participants. However, to the extent that participants are already skilled at recognizing White faces due to the societal status and representation of White people, there may not have been as large of an improvement as expected for Black faces (i.e., a ceiling effect).

Method

Study Design

This experiment was a 2x2x2 mixed design, including within-subjects factors of stimulus race (Black, White) and time of testing (before manipulation, after manipulation), and a between-subjects factor of experimental condition (holistic training, matched control).

Participants

I recruited undergraduate participants through the SONA Systems website managed by the University of California, Riverside Psychology department.. As compensation, participants received 1 research credit for an expected 1 hour of participation. An a priori power analysis indicated that, with 48 observations per person, I would have 80% power to capture an effect size of at least .14 with $N = 112$ participants. To be conservative, I doubled this amount and aimed to recruit $N = 224$ participants. I made my study available on SONA for two weeks, which resulted in an initial sample size of $N = 364$.

After removing individuals who did not pass the attention check (i.e., “If you are reading this carefully, select ‘Three’ from the options below”; $n = 5$), had incomplete data ($n = 27$), or opted to exclude their data from analyses ($n = 12$), the final sample consisted of 320 participants ($M_{\text{age}} = 19.46$, $SD = 1.76$). Most participants were female ($n = 191$) and identified as Asian ($n = 115$) or Latinx ($n = 134$). Of this final sample, 151 completed the holistic training condition and 169 individuals completed the control condition.

Materials and Manipulation

Face Stimuli. For the recognition memory tasks and experimental conditions in Experiments 1-3, I used Black and White male faces from the Chicago Face Database (Ma et al., 2015).

I selected 44 faces for the holistic training and matched control conditions, most in groups of two or three based on perceived similarity in appearance. I then divided the remaining number of Black and White male faces in the face database by the maximum number of recognition memory tasks participants would take in a given experiment (i.e., four recognition memory tasks in Experiment 2), which resulted in 12 faces of each race per task. Therefore, the pre-manipulation and post-manipulation experimental conditions in Experiment 1 each contained 24 male faces (12 Black, 12 White).

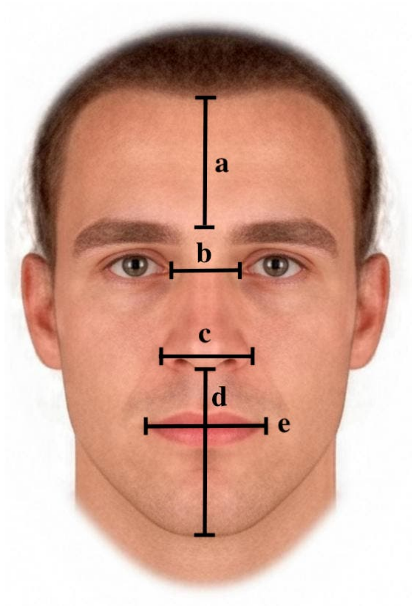
Face Editing. I used the free online image editing software Photopea to alter feature measurements and appearance of face stimuli displayed in the training and matched control conditions.

Experimental Manipulation. Participants completed one of two experimental conditions, the holistic training or a matched control. Pretesting indicated that both conditions required approximately 30-40 minutes to complete.

Holistic Training. In developing the holistic training, my goals were to 1) direct participants' attention to variability in the appearance of and spatial relationships between features, and 2) provide practice for processing faces in this way. I designed the training to focus on five facial features with substantial variability across ethnic groups (Fang et al., 2011): a) forehead length, b) inner-eye space, c) nose width, d) nose-to-chin length, and e) mouth width (FINNM, for short; Figure 1). Given my intention to improve face recognition broadly, I selected these features with the expectation that the skills participants learn could be applied to faces from a wide range of ethnic backgrounds; however, because a large portion of the face recognition literature compares memory for Black versus White faces, I created the training with Black and White male stimuli.

Figure 1

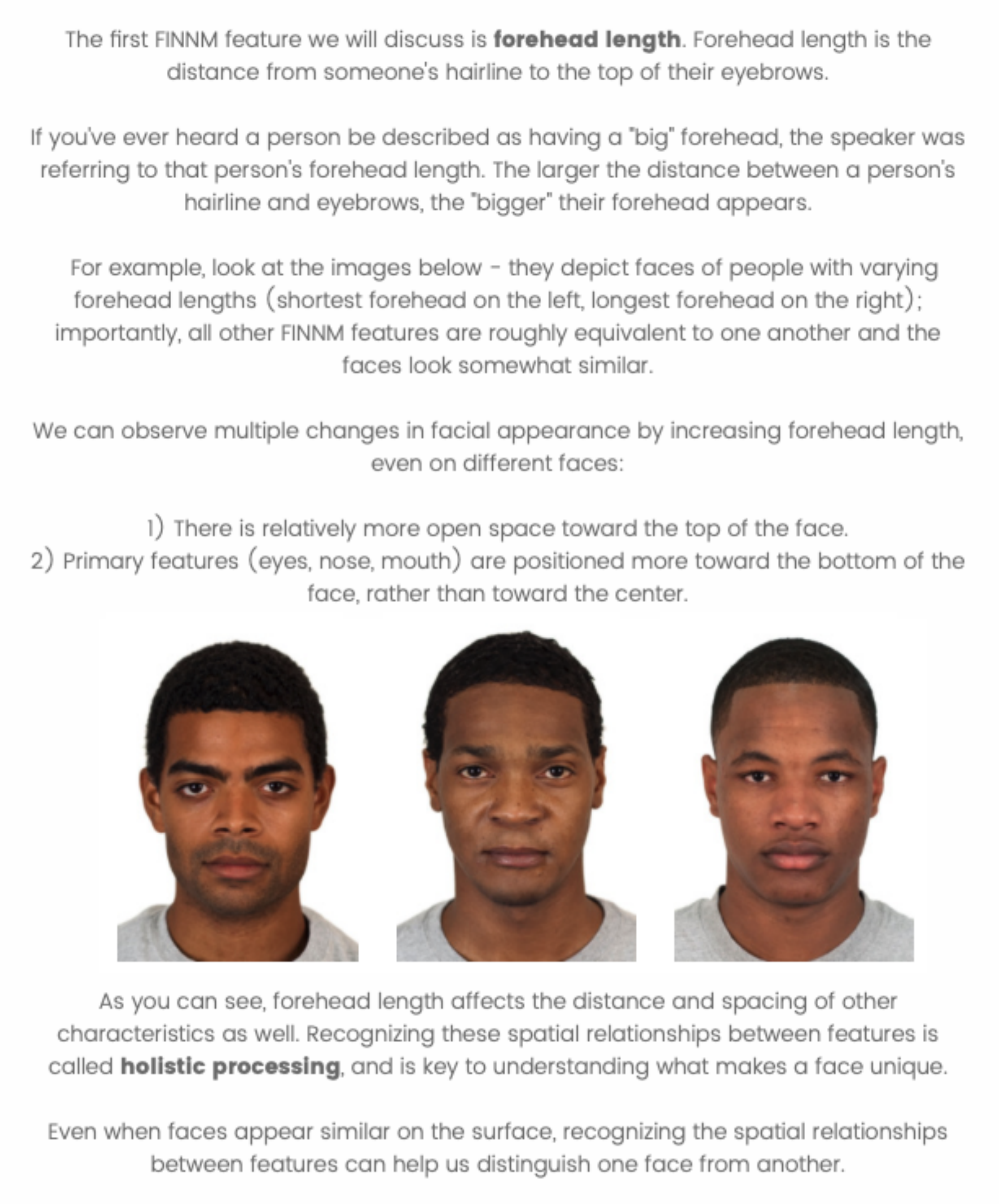
FINNM Training Features



Note. Facial features of focus in the holistic processing training: a) forehead length, b) inner-eye space, c) nose width, d) nose-to-chin length, and e) mouth width. Referred to in shorthand as the "FINNM" features.

The training consisted of two modules: one learning module with lessons and quizzes for each FINNM feature, and a second module synthesizing the contents of the learning module with a focus on holistic processing practice. To maximize the effectiveness of the training, I incorporated two evidence-based instruction techniques known to improve learning outcomes. The first technique, active learning, is when learners engage with instructional material in a reflective, hands-on fashion (Felder & Brent, 2009). I incorporated active learning into the training by providing participants with writing prompts about faces depicted, as well as activities that involved arranging faces in order based on a given FINNM feature (e.g., from smallest to largest nose). The second technique, interleaving, involves displaying related concepts in a mixed fashion to facilitate deeper learning (Pan, 2015). I incorporated interleaving by integrating holistic processing practice throughout Module 1, rather than reserving all holistic processing trials for Module 2.

Module 1 contained lessons that introduced participants to each FINNM feature and how variability within these features can impact the appearance of a face. It also introduced participants to the concept of holistic processing and its role in accurate face recognition. Each lesson began with a definition of the target feature, then modeled how to recognize variability in the feature and its influence on the configuration and appearance of a face. In this demonstration, participants observed three distinct yet visually similar faces that varied on the lesson's target feature (e.g., short, medium, and long forehead lengths), while the other FINNM features were matched in size (Figure 2). I accomplished this by digitally modifying the faces, such that the lesson feature measurements differed by more than 20 pixels from one face to the next, while all other FINNM features measured within 20 pixels across the three faces. With ecological validity in mind, I opted to use different faces for the lessons to mirror the experience of encountering different people in daily life. Once I finished editing the faces, I pilot tested them among a group of perceivers to ensure they appeared natural.



The first FINNM feature we will discuss is **forehead length**. Forehead length is the distance from someone's hairline to the top of their eyebrows.

If you've ever heard a person be described as having a "big" forehead, the speaker was referring to that person's forehead length. The larger the distance between a person's hairline and eyebrows, the "bigger" their forehead appears.

For example, look at the images below - they depict faces of people with varying forehead lengths (shortest forehead on the left, longest forehead on the right); importantly, all other FINNM features are roughly equivalent to one another and the faces look somewhat similar.

We can observe multiple changes in facial appearance by increasing forehead length, even on different faces:

1) There is relatively more open space toward the top of the face.
2) Primary features (eyes, nose, mouth) are positioned more toward the bottom of the face, rather than toward the center.

As you can see, forehead length affects the distance and spacing of other characteristics as well. Recognizing these spatial relationships between features is called **holistic processing**, and is key to understanding what makes a face unique.

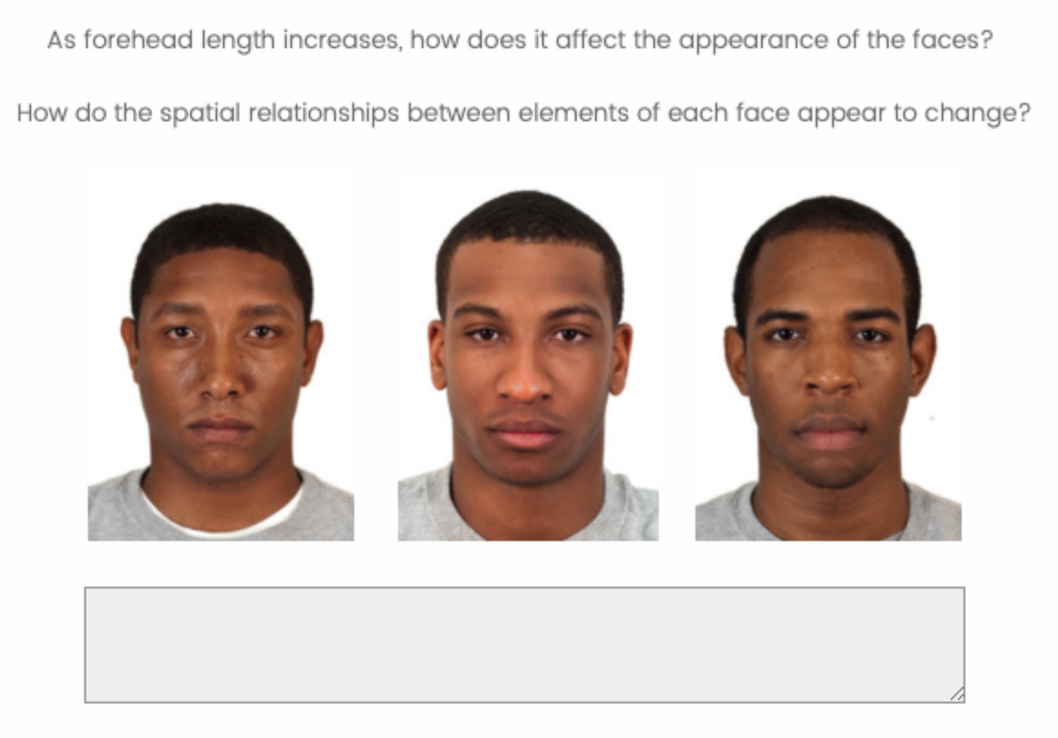
Even when faces appear similar on the surface, recognizing the spatial relationships between features can help us distinguish one face from another.

Figure 2

Module 1 Lesson Introduction

Note. The Module 1 lesson introduction for forehead length. Lesson introductions for the remaining FINNM features were similarly structured.

After an introduction to the lesson feature, Module 1 lessons continued with four trials depicting groups of three faces, one group per trial, that varied on the target feature while matched on the other FINNM features. In these trials, participants were prompted to reflect on and write about how variability in the target feature affected each face's appearance, specifically in terms of how it influenced the spatial configuration between features (Figure 3). To ensure participants devoted an adequate amount of time to the task, the button to proceed to the next trial appeared once 20 seconds had passed.



As forehead length increases, how does it affect the appearance of the faces?

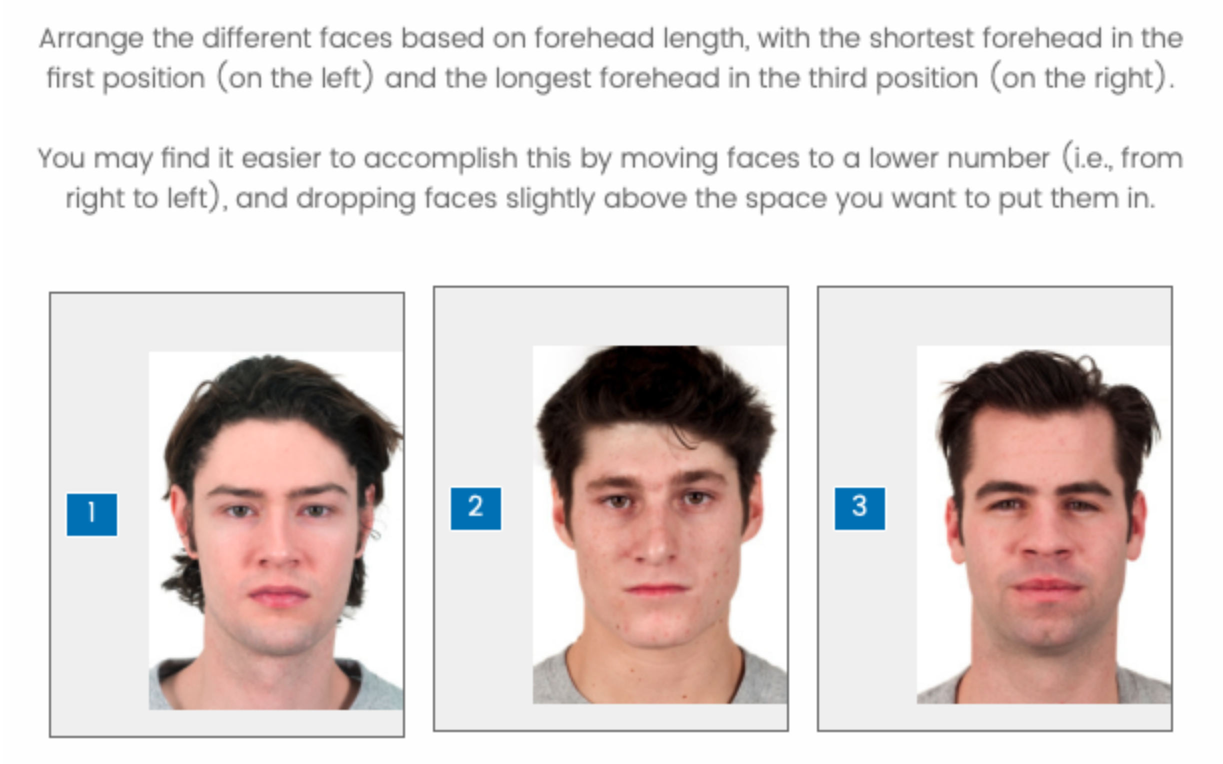
How do the spatial relationships between elements of each face appear to change?

Figure 3

Module 1 Lesson Reflection Trial

Note. A reflection trial from the Module 1 forehead length lesson.

After the writing prompt trials, participants completed three trials in which they arranged three faces – also varying on the lesson feature yet matched on the other FINNM features – in order based on the lesson feature (e.g., from shortest to longest forehead; Figure 4).



Arrange the different faces based on forehead length, with the shortest forehead in the first position (on the left) and the longest forehead in the third position (on the right).

You may find it easier to accomplish this by moving faces to a lower number (i.e., from right to left), and dropping faces slightly above the space you want to put them in.

Figure 4

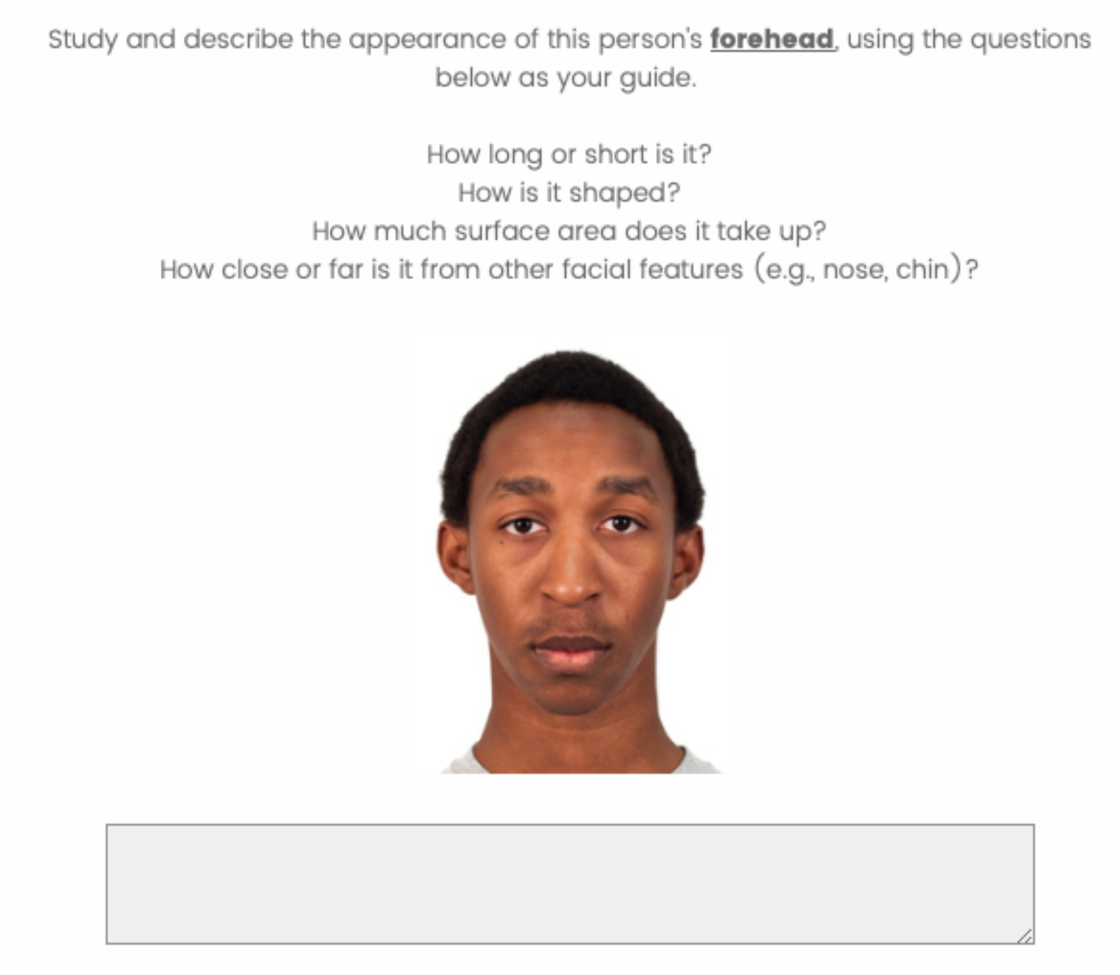
Module 1 Lesson Ordering Trial

Note. An ordering trial from the Module 1 forehead length lesson. For this trial to be correct, the participant would have switched faces 1 and 2.

In the final trial of each lesson, participants engaged in a holistic processing exercise similar to what they would complete in Module 2. Specifically, participants observed a face and were tasked with describing the lesson feature in relationship to other elements of that face (Figure 5). To ensure participants devoted an adequate amount of time to the task, the button to proceed to the next trial appeared once 20 seconds had passed.

Figure 5

Module 1 Lesson Holistic Processing Trial

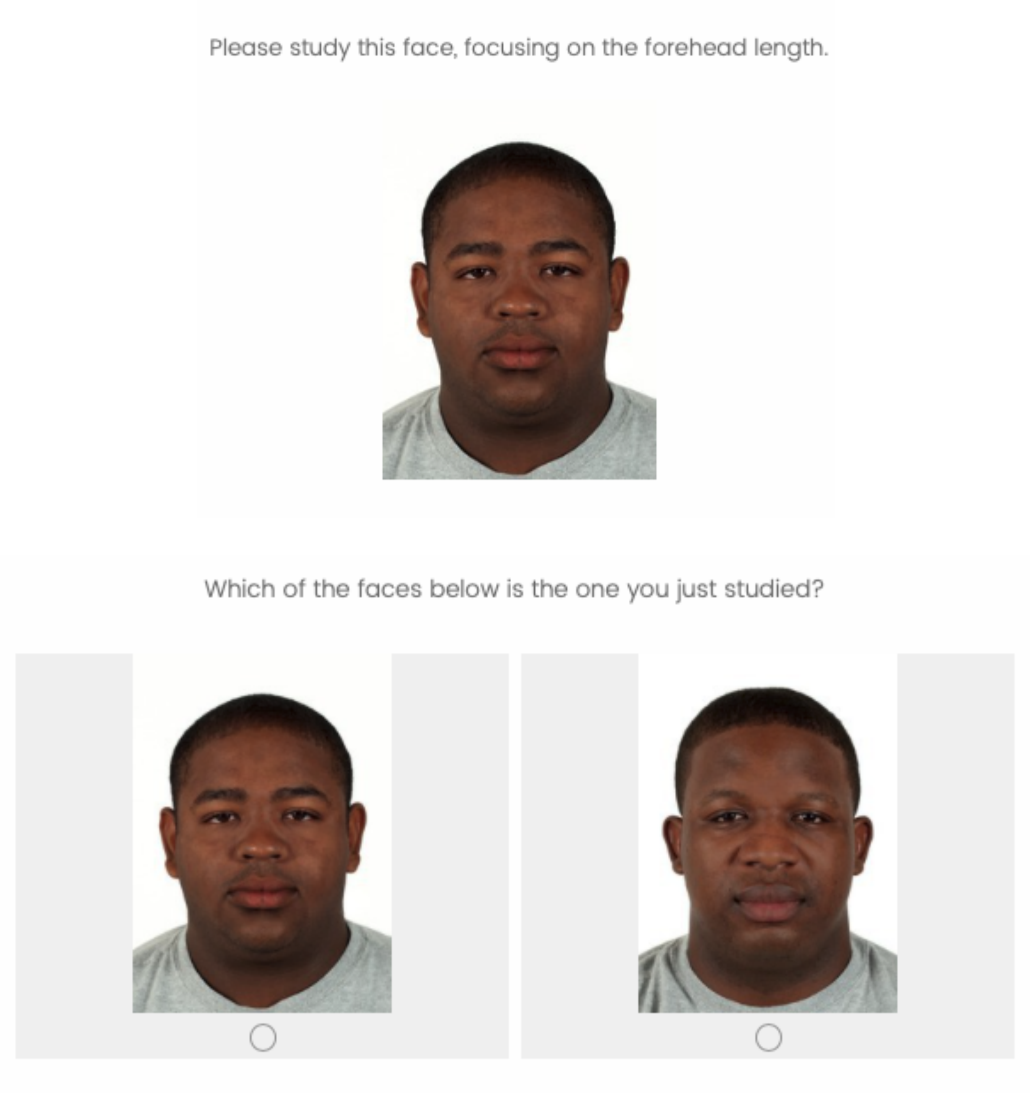


Study and describe the appearance of this person's **forehead**, using the questions below as your guide.

How long or short is it?
How is it shaped?
How much surface area does it take up?
How close or far is it from other facial features (e.g., nose, chin)?

Note. A holistic processing trial within the Module 1 forehead length lesson.

Each lesson of Module 1 was followed by a quiz pertaining to the lesson feature. In the quizzes, participants completed four trials in which they viewed a target face for two seconds, focusing particularly on the lesson feature. The target face then disappeared and two new faces appeared, a distractor face and the target face once again. The faces were similar in appearance and their FINNM feature measurements were matched (i.e., within 20 pixels) except for the target feature, which was distinctly different between the faces (i.e., greater than 20 pixels). In each of these trials, participants chose which of the two faces they had just seen (Figure 6).



Please study this face, focusing on the forehead length.

Which of the faces below is the one you just studied?

○ ○

Figure 6

Module 1 Quiz Trial

Note. A trial from the Module 1 forehead length quiz. Participants viewed the top panel for two seconds, then it disappeared and the bottom panel was displayed.

My intention for Module 2 was to synthesize the lessons in Module 1, providing participants with further opportunity to practice attending to the FINNM features holistically. Module 2 began by demonstrating the insights that could be gleaned from holistic processing of the FINNM features (Figure 7).

At the end of each lesson, you have studied and described the appearance of a face paying specific attention to one FINNM feature at a time.

Now that you are familiar with the features, in this final module you will practice holistic processing by 1) observing the shape and appearance of features, and 2) their relationships with one another collectively.
In other words, when engaging in holistic processing you are considering features in the context of their surroundings, or in relation to other features.

To provide an example of this, let's return to the scenario with your new coworker, Anthony, from the beginning of this training.

If you engaged in holistic processing of the FINNM features during this initial meeting, you might notice a few of the following details about his face:

- His forehead length is average and is a consistent length from one side to the other, giving his forehead a rectangular shape
- His forehead is wider than his chin, creating a face shape similar to a guitar pick
- His forehead is far from his chin, conveying that his face is long
- He has an average amount of inner-eye space, with almond-shaped eyes that aren't very close together nor very far apart
- The outside of his eyes are fixed near his temples
- His eyes are far from his mouth, so a lot of surface area is dedicated to his cheeks. He high cheekbones right below his eyes
- He has a slightly wide nose with a thin bridge
- His nose-to-chin area is decently long, and his mouth is positioned about halfway down (i.e., his nose-to-mouth and mouth-to-chin are roughly equal distances)
- He has a wide mouth with full lips that extend out near the edges of his face

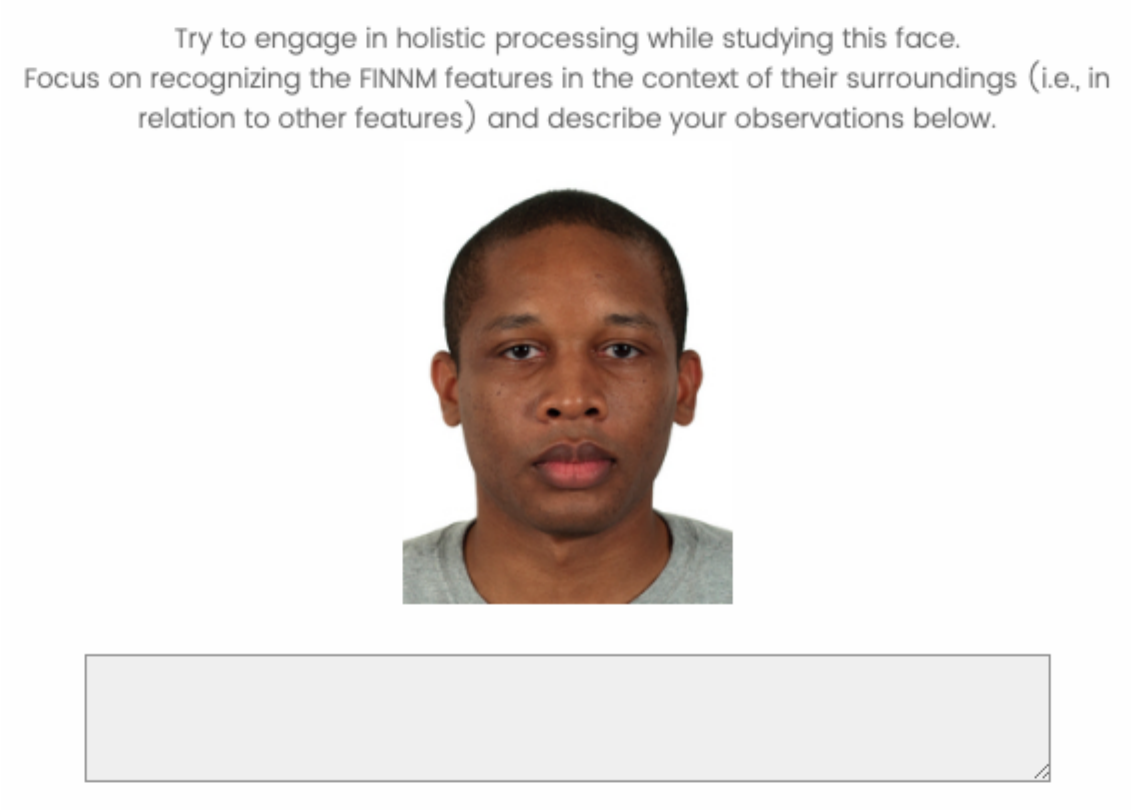
Though you likely wouldn't recognize all of these characteristics, with continued practice, you can develop a handful of go-to characteristics that you *would* be able to recognize in a brief encounter.

Figure 7

Module 2 Introduction

Note. Introduction to the final module of the training.

Participants then completed four trials in which they observed a face and described how the FINNM features related to other features and characteristics of the face (Figure 8). To ensure participants devoted an adequate amount of time to the task, the button to proceed to the next trial appeared once 20 seconds had passed.



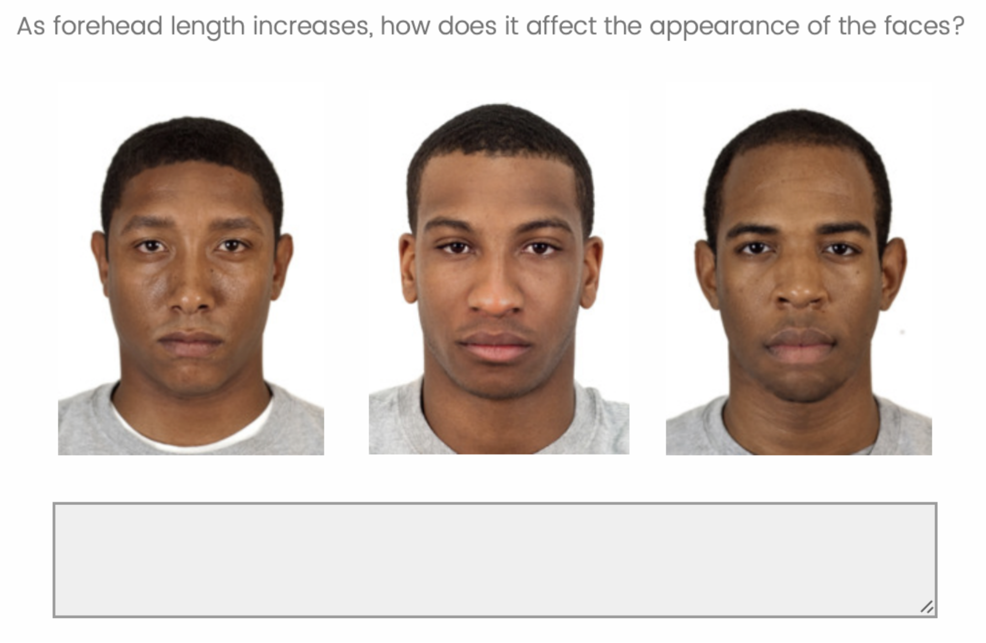
Try to engage in holistic processing while studying this face.
Focus on recognizing the FINNM features in the context of their surroundings (i.e., in relation to other features) and describe your observations below.

Figure 8

Module 2 Trial

Note. A holistic processing trial within Module 2.

Matched Control. The matched control condition mirrored the holistic training with one important distinction: rather than directing participants' attention to the FINNM features and how they influence the spatial configuration among elements of a face, the matched control condition focused solely on how the FINNM features affect facial appearance generally (Figure 9). Further, the training introduced participants to the concept of holistic processing and highlighted its importance for accurate face recognition, which was absent from the matched control condition.



As forehead length increases, how does it affect the appearance of the faces?

Figure 9

Module 1 Matched Control Reflection Trial

Note. A reflection trial from the Module 1 forehead length matched control condition. Unlike the training condition, it did not contain the additional prompt, "How do the spatial relationships between elements of the face seem to change?"

Procedure

After providing consent, participants began the experiment by completing a baseline recognition memory task that consisted of three components: learning, filler, and recognition. In the initial learning task, participants received instructions to pay attention to the faces presented because they would be asked about the faces at a later time. They then passively viewed 12 male faces (6 Black, 6 White) one at a time, in a random order for two seconds each. Next, participants completed a filler task intended to clear the contents of the learning task from their working memory. The filler task consisted of a short passage about how Cheerios are made, with three reading comprehension questions based on the passage. In the subsequent recognition task, participants viewed all faces from the learning task, as well as 12 novel male faces (6 Black, 6 White), in a random order. For each face, participants responded to the question, "Was this image presented before?," and click "yes" or "no".

Participants then completed the holistic training or matched control condition, which was randomly assigned within the survey. After the experimental manipulation, participants completed the recognition memory task again with a new set of faces. They then reported how much effort they put into the study and whether they had seen any of the faces previously. Finally, participants completed a demographics questionnaire and were debriefed.

Results

Comparing Performance on Experimental Condition Tasks

I began analyses by exploratorily comparing accuracy on lesson ordering trials, as well as quizzes, between the two conditions. A lesson ordering trial was correct if participants arranged the three faces in order based on the lesson feature (e.g., smallest to largest nose width), whereas a quiz trial was correct if participants selected the face that was displayed previously and not the distractor face.

A Welch two samples *t*-test indicated that there was no reliable difference in accuracy on lesson ordering trials between training ($M = 0.80$, $SD = 0.10$) and control ($M = 0.79$, $SD = 0.14$) participants, $t(313.62) = 1.36$, $p = .17$, $d = .15$. However, training participants ($M = 0.98$, $SD = 0.03$) were reliably more accurate on the quizzes than control participants ($M = 0.97$, $SD = 0.08$), albeit a small effect, $t(243.39) = 2.37$, $p = .02$, $d = .25$.

Change in Recognition Memory by Condition

From trial-level "yes, no" responses on the two recognition memory tests, I tallied participants' number of hits, false alarms, misses, and correction rejections for Black and White faces both before and after the manipulation. I then computed each participant's hit and false alarm rate (i.e., number of hits/false alarms divided by the number of possible hits/false alarms) for the two memory tests, and adjusted for rates of 0 or 1 by adding or subtracting 0.5. Within the two tests, hit rates of 0 occurred once and 4 times, false alarm rates of 0 occurred 195 and 215 times, hit rates of 1 occurred 185 and 145 times, and false alarm rates of 1 occurred once and twice. Descriptive statistics for hits and false alarm rates can be found in Appendix B.

Finally, I computed each participant's average memory strength (operationalized as signal detection parameter *d*') for Black and White faces during both tests. The number of participants with *d*' below zero across trials, indicating chance performance throughout the experiment, can be found in Appendix A.

Welch two samples *t*-tests indicated that training and control participants' *d'* did not meaningfully differ in their baseline recognition of Black faces, *t*(313.62) = 1.25, *p* = .21, *d* = .14, nor of White faces, *t*(317.95) = -0.42, *p* = .68, *d* = .05. Prior to the manipulation, training participants recognized White faces (*M* = 2.16, *SD* = 0.89) descriptively better than Black faces (*M* = 2.07, *SD* = 0.94), but the difference was not reliable, *t*(150) = 1.14, *p* = .26, *d* = .11. Control participants, on the other hand, recognized White faces (*M* = 2.21, *SD* = 0.98) reliably better than Black faces (*M* = 1.93, *SD* = 0.93) prior to the manipulation, *t*(168) = 3.12, *p* = .002, *d* = .28.

After the manipulation, training participants recognized White faces (*M* = 2.15, *SD* = 1.11) reliably better than Black faces (*M* = 1.84, *SD* = 1.07), *t*(150) = 3.70, *p* < .001, *d* = .28. Control participants also recognized White faces (*M* = 2.19, *SD* = 1.03) reliably better than Black faces (*M* = 1.77, *SD* = 1.22) as well, *t*(168) = 5.26, *p* < .001, *d* = .38.

I then computed each condition's average change in recognition memory for Black and White faces. Contrary to predictions, training participants (*M* = -0.22, *SD* = 1.41) did not meaningfully differ from control participants (*M* = -0.17, *SD* = 1.56) in their recognition memory change for Black faces, *t*(317.96) = -0.34, *p* = .73, *d* = .04; in fact, both conditions displayed poorer memory for Black faces after versus before the manipulation, with training participants descriptively demonstrating a greater decline than control participants. Similarly, training participants (*M* = -0.01, *SD* = 1.35) were nearly identical to control participants (*M* = -0.01, *SD* = 1.37) in their recognition memory change for White faces, *t*(314.78) = -0.005, *p* = .99, *d* = .001.

Primary Analyses

There were three levels of nesting within the data for this experiment: stimuli (two Level 1 units, Black and White) within time points (two Level 2 units, pre- and post-manipulation) within participants (320 Level 3 clusters). Each Level 3 cluster (i.e., participant) contained the same number of Level 1 and Level 2 units.

I first examined the intraclass correlation (ICC) of the unconditional model to determine whether a multilevel model would be appropriate. The ICC was .49, indicating that there was sufficient variability at Level 3 to justify the use of a multilevel model. After preparing contrast codes for the predictors, I specified a multilevel model (Table 1) predicting recognition memory (*d'*) from a three-way interaction between stimulus race (White = -1, Black = 1; Level 1) time of test (pre-manipulation = -1, post-manipulation = 1; Level 2), and condition (matched control = -1, holistic training = 1; Level 3). I entered all predictors as fixed effects and included participants as a random intercept. Due to contrast coding, coefficients in the model reflect mean differences between the reference group and comparison group for each predictor.

All models conducted were fit using restricted maximum likelihood estimation with the *lme4* package (Bates et al., 2015) in R version 4.3.2 (R Core Team, 2023). Inferential tests of the fixed effects used the Satterthwaite method of degrees of freedom approximation, computed with the *lmerTest* package (Kuznetsova et al., 2017).

Table 1

Experiment 1 Model of Recognition Memory Strength from Time of Test, Stimulus Race, and Condition

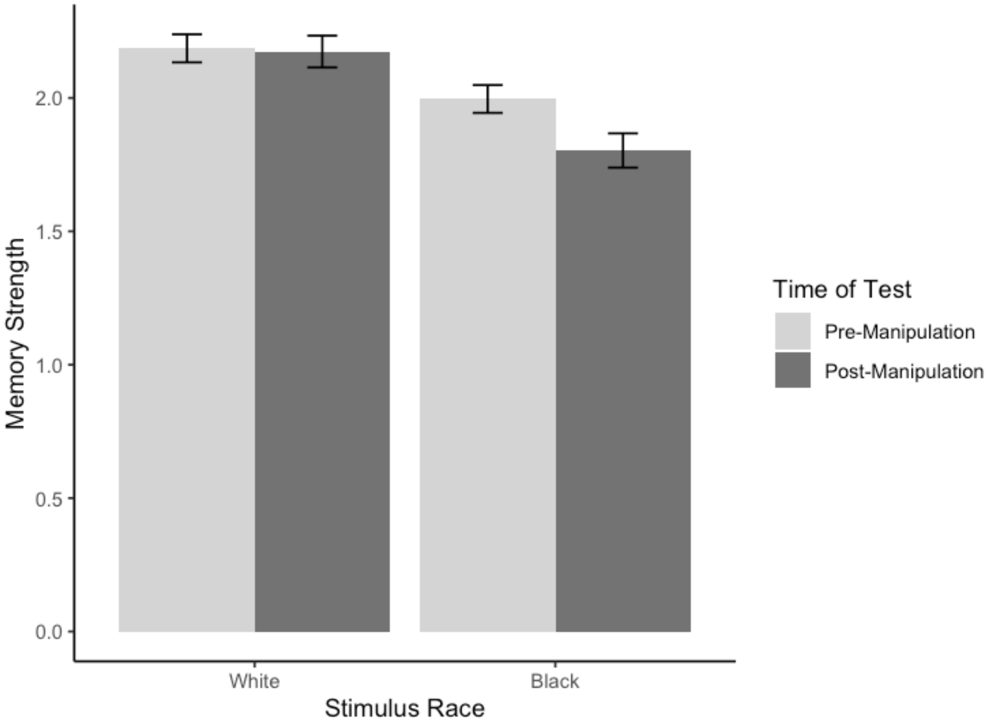
	Estimate	Standard Error
(Intercept)	2.041***	0.041
Time of Test	-0.052*	0.023
Stimulus Race	-0.138***	0.023
Condition	0.015	0.041
Time of Test × Stimulus Race	-0.046*	0.023
Time of Test × Condition	-0.007	0.023
Stimulus Race × Condition	0.037	0.023
Time of Test × Stimulus Race x Condition	-0.007	0.023
AIC	3571.1	
BIC	3622.7	
ICC	0.4	

Note. *$p < .05$, ***$p < .001$. Time of test reflects pre-manipulation versus post-manipulation, such that positive values represent better recognition after the manipulation. Stimulus race reflects White versus Black faces, such that positive values represent better recognition of Black faces. Condition reflects matched control versus holistic training, such that positive values represent better recognition within the holistic training condition.

Main effects of stimulus race and time of test were qualified by an interaction between the two variables, such that the difference in recognition before versus after the manipulation varied depending on stimulus race, *t*(954) = -1.97, *p* = .05, 95% CI = [-0.09, -0.003] (Figure 10). Pairwise comparisons of estimated marginal means indicated that participants' average recognition of Black faces was reliably worse after the manipulation (*M* = 1.81, *SE* = 0.06) compared to before (*M* = 2.00, *SE* = 0.06), *t*(954) = -2.97, *p* = .003; there was no difference in recognition of White faces before versus after the manipulation.

Figure 10

Experiment 1 Stimulus Race by Time of Test Interaction



Note. Average recognition memory strength for Black and White faces, measured before and after the manipulation. Error bars represent standard errors.

Discussion

My primary hypothesis in Experiment 1 was that training participants would display a more pronounced improvement in recognition memory from pre- to post-manipulation than control participants; this would suggest that the training imparted unique knowledge that aided participants in distinguishing between faces, to a greater extent than any influence the control condition may have had. Instead, training and control participants did not differ in their degree of recognition change from pre- to post-manipulation: recognition of White faces remained approximately the same, and recognition of Black faces *declined* (opposite to my expectations) in both conditions.

Prior to the manipulation, I anticipated that participants in both conditions would better recognize White than Black faces on average, which did occur. The current sample consisted primarily of Asian and Latinx participants, such that all faces viewed during the study were racial outgroup members to most participants; this finding aligns with my own prior work showing that American participants consistently better recognize White than Black faces, even when both groups are racial outgroups to the participants (Simon et al., 2023).

After the manipulation, it was an open question how recognition of White faces might change. Both conditions could plausibly display improved recognition, due to the holistic processing skills learned and practiced among those in the training condition, or due to the increased attention to specific facial features among those in the control condition; however, it was also possible that recognition might not alter very much (i.e., a ceiling effect) if participants were already skilled at recognizing White faces due to the societal status and representation of White people. The lack of change in both conditions' average White recognition from pre- to post-manipulation supports this explanation.

Further, I anticipated that Black recognition would improve to a greater extent post-manipulation in the training condition relative to the control condition. To my surprise, neither condition displayed an improvement in recognition for Black faces; instead, both conditions displayed worse recognition for Black faces, and were similar in their average amount of decline. Though this result was contrary to my initial expectations, it aligns with the outcome of some other face recognition interventions (Malpass, 1981; Woodhead et al., 1979).

Finally, it is worth noting that the two conditions were nearly identical, on average, in their recognition of Black and White faces after the manipulation. This begs the question of whether the training and control conditions are sufficiently distinct from one another; Experiment 2 will be able to offer further insight into this possibility.

Experiment 2

In Experiment 2 of this research, I sought to explore whether my intervention had a lasting influence on face recognition. It largely replicated Experiment 1, with the addition of two additional recognition memory tasks occurring 24 hours after and one week after the initial experimental session. If my training was effective in teaching how to distinguish between faces, participants who completed the training would demonstrate an increased ability to recognize faces after the manipulation and presumably at later points in time. However, there were multiple possible outcomes for how training participants' recognition memory would evolve: (1) recognition could continue to improve with time as participants accrued more experience applying the training material in their daily lives; (2) participants could maintain an elevated level of recognition comparable to their initial post-manipulation score; or (3) recognition could decline if their learning faded over time.

To begin examining the impact of my training on recognition of same- versus other-race faces, I recruited non-Latinx, White identifying participants in the United States for this experiment. Cross-race effect literature has consistently shown that White individuals display substantially better recognition for same- than other-race faces (Meissner & Brigham, 2001). However, after completing a face recognition intervention, White participants have displayed increased similarity in their recognition of same- and other-race faces, with poorer recognition for same-race faces and improved recognition for other-race faces, on average (Hugenberg et al., 2007; Malpass et al., 1973). That said, it was interesting to consider how White participants' recognition for same- versus other-race faces might change as a result of this training: though they may display a similar degree of improvement for both races post-manipulation, if they were already skilled at recognizing same-race faces at baseline, they may not display an improvement for same-race faces (i.e., a ceiling effect) as they might for other-race faces, aligning with previous findings.

The final addition to this experiment involved exploring the potential role of memory self-efficacy in face recognition. Memory self-efficacy refers to a person's evaluation of their own memory abilities (Beaudoin & Desrichard, 2011). It is conceptualized in two ways: as a person's assessment of their prospective performance on a specific memory task (Berry & West, 1993; Hertzog et al., 1990), or as a person's general beliefs about their memory abilities, either globally or within a specific domain (e.g., remembering names; Hertzog & Dixon, 1994). In the context of face recognition, some work has examined participants' prospective confidence in their ability to perform on face recognition tasks, finding that participants are more confident in recognizing same- than other-race faces within these tasks (Hourihan et al., 2012; Smith et al., 2004); however, I sought to explore how the general beliefs people hold about their memory abilities, not in the context of a specific task, may relate to face recognition.

The literature on aging suggests that participants' negative beliefs about their ability to remember faces may adversely affect their performance. Older adults who reported lower memory self-efficacy performed worse on memory-oriented tasks than older adults with high memory self-efficacy (Beaudoin & Desrichard, 2017; Desrichard & Köpetz, 2005); further, older adults who completed an intervention designed to boost memory self-efficacy demonstrated improved memory performance after the intervention (West et al., 2008).

I anticipated that self-efficacy would positively correlate with memory performance, such that the higher (or lower) a person's memory self-efficacy was at a given time point, the better (or worse) their face recognition would be at that time point as well. It was an exploratory question whether the two conditions would differ on average in their degree of memory self-efficacy. If training participants perceived their condition to be more helpful for face recognition than control participants did, training participants may report greater memory self-efficacy than their control condition counterparts. However, given that both conditions would receive learning material about faces, participants may report similar degrees of memory self-efficacy regardless of the condition they were assigned to.

Method

Study Design

This experiment was a 2x2x2 mixed design, including within-subjects factors of stimulus race (Black, White) and time of testing (before manipulation, after manipulation, 24 hours later, 1 week later), and a between-subjects factor of condition (holistic training, matched control).

Participants

I recruited non-Latinx, White adults residing in the United States through the CloudResearch Connect platform. As compensation, participants received $15 for an expected 1 hour of participation in the initial testing session, and $3.75 each for an expected 15 minutes of participation in the two subsequent testing sessions. An a priori power analysis indicated that, with 96 observations per person, I would have 80% power to capture an effect size of at least .14 with $N = 97$ participants. To be conservative, and to account for attrition, I aimed to recruit at least 2.25x this amount (i.e., $N = 224$ participants).

Removing individuals who did not pass the quality control or attention checks ($n = 8$), did not return for all time points ($n = 15$), had incomplete data across the testing sessions (*n* = 21), or did not identify as exclusively White (*n* = 4) resulted in a final sample of 196 participants (M_{age} = 39.73, *SD* = 10.59; 106 males). Of this final sample, 94 completed the holistic training condition and 102 individuals completed the control condition.

Materials, Manipulation, and Measures

Face Stimuli. I used the Experiment 1 Chicago Face Database stimuli for the experimental conditions and pre- and post-manipulation recognition memory tasks. For the two additional recognition memory tasks in Experiment 2, I selected 24 faces per task (12 Black, 12 White) from the Black and White male faces that remained from the Chicago Face Database.

Experimental Manipulation. As in Experiment 1, participants completed the holistic training or matched control conditions.

Memory Self-Efficacy. To measure participants' self-efficacy with regard to remembering faces, I selected and adapted five items from the Metamemory in Adulthood Capacity subscale (Dixon & Hultsch, 1983). On a sliding scale from 0 (completely disagree) to 10 (completely agree), participants indicated their agreement with beliefs about their face recognition abilities (e.g., "I am able to recognize someone I've met once before"; αs = .91-.93 for time points 1-3). Higher scores represent greater self-efficacy.

Procedure

Data collection for this experiment took place during three separate time points over the course of a week. After providing informed consent, participants began the initial testing session by completing two quality control checks; only individuals who correctly responded to these checks were able to continue. The remainder of the initial testing session was largely identical to Experiment 1: participants completed a baseline recognition memory task, followed by the randomly-assigned training or control condition, then concluded with a second recognition memory task. After the post-manipulation memory task, participants completed the memory self-efficacy questionnaire, answered demographic questions, and received instructions about the subsequent time points. The instructions communicated that the second and third surveys would be available for completion in approximately 24 hours and 1 week, and that participants would have 48 hours to complete the surveys once they became available.

Noting that the majority of participants completed the initial testing session at around 9 AM, I released the second survey at 9 AM the following day. Participants received notification at this time that the second survey was available and that they would have 48 hours to complete it. During this second testing session, participants engaged in another recognition memory task, completed the memory self-efficacy questionnaire once again, and read instructions pertaining to the final survey.

One week after posting the initial survey, I released the final survey at 9 AM for participants to complete. Once again, participants received notification at this time that the final survey was available and that they would have 48 hours to complete it. The survey was nearly identical to the second survey, consisting of another recognition memory task and the memory self-efficacy questionnaire. After completing these components, participants were fully debriefed.

Results

Comparing Performance on Experimental Condition Tasks

As in Experiment 1, I began analyses by exploratorily comparing accuracy on lesson ordering trials, as well as quizzes, between the two conditions. Once again, a lesson ordering trial was correct if participants arranged the three faces in order based on the lesson feature (e.g., smallest to largest nose width), whereas a quiz trial was correct if participants selected the face that was displayed previously and not the distractor face.

A Welch two samples *t*-test indicated that there was no reliable difference in accuracy on lesson ordering trials between training (*M* = 0.77, *SD* = 0.09) and control (*M* = 0.74, *SD* = 0.12) participants, *t*(190.67) = 1.43, *p* = .15, *d* = .20. Similarly, there was no reliable difference in quiz accuracy between training (*M* = 0.97, *SD* = 0.03) and control (*M* = 0.96, *SD* = 0.07) participants, *t*(138.56) = 1.17, *p* = .24, *d* = .16.

Change in Recognition Memory by Condition

From trial-level "yes, no" responses on the four recognition memory tests, I tallied participants' number of hits, false alarms, misses, and correction rejections for Black and White faces. I then computed each participant's hit and false alarm rate (i.e., number of hits/false alarms divided by the number of possible hits/false alarms) for each memory test, and adjusted for rates of 0 or 1 by adding or subtracting 0.5. Across the four memory tests, hit rates of 0 ranged from 0-2 instances, false alarm rates of 0 ranged from 107-142 instances, hit rates of 1 ranged from 81-90 instances, and false alarm rates of 1 ranged from 0-2 instances. Descriptive statistics for hits and false alarm rates can be found in Appendix C.

Finally, I computed each participant's average recognition memory strength (operationalized as signal detection parameter *d*') for Black and White faces during all tests. Descriptive statistics of White and Black recognition for both conditions during the tests can be found in Table 2.

Table 2

Experiment 2 Descriptive Statistics

	Holistic Training				**Matched Control**			
	Black Faces		**White Faces**		**Black Faces**		**White Faces**	
Time of Test	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Before Manipulation	1.89^{a}	1.05	2.28^{a}	0.99	1.61^{e}	1.01	$1.87^{e,h}$	1.04
After Manipulation	1.76^{b}	1.03	2.40^{b}	0.97	1.73^{f}	1.13	$2.15^{f,h,j}$	1.09
24 Hours Later	1.93^{c}	1.18	2.24^{c}	1.03	1.87^{k}	1.06	1.94^{j}	1.01
1 Week Later	1.93^{d}	1.03	2.25^{d}	1.03	$1.61^{g,k}$	1.11	1.89^{g}	1.14

Note. Means with shared superscripts are reliably different from one another. Shared superscripts in the same row indicate a reliable difference between training and control conditions at the same time point. Shared superscripts in the same column indicate a reliable difference between time points within a single condition.

Welch two samples *t*-tests indicated that training and control participants meaningfully differed in their baseline recognition of White faces, $t(193.76) = 2.87$, $p = .005$, $d = .41$, and descriptively differed in their recognition of Black faces, $t(190.99) =$

1.87, $p = .06$, $d = .27$. This indicates that training participants were better, on average, at recognizing faces than control participants from the start, which was not intended.

Within-subjects *t*-tests indicated that training participants recognized White faces reliably better than Black faces at all time points, *t*'s < 6.85, *p*'s < .002, *d*'s < .63. Alternatively, control participants recognized White faces reliably better than Black faces at most time points, *t*'s < 4.59, *p*'s < .03, *d*'s < .38; the exception to this was at Time 3, during which control participants descriptively better recognized White than Black faces, $t(101) = 0.74$, $p = .46$, $d = .06$.

Primary Analyses

There were three levels of nesting within the data for this experiment: stimuli (two Level 1 units, Black and White) within time points (four Level 2 units, pre-manipulation, post-manipulation, 24 hours later, and 1 week later) within participants (196 Level 3 clusters). Each Level 3 cluster (i.e., participant) contained the same number of Level 1 and Level 2 units. The intraclass correlation (ICC) of the unconditional model was .45, indicating that there was sufficient variability at Level 3 to justify the use of multilevel modeling.

In multilevel modeling, how categorical predictors are specified (e.g., how levels of the predictor are coded) has notable implications for the model's interpretation; though the coding system does not alter the omnibus model or its results, the regression coefficients will have different interpretations and answer different research questions (Yaremych et al., 2021). I employed contrast coding because I was interested in tracking average recognition change from one measurement time point to the next. Contrast coding permits greater flexibility in the comparisons one makes, which allowed me to sequentially compare each measurement to the time point that followed (i.e., specify different reference levels for each comparison).

In this analysis, I used three contrasts to compare participants' recognition memory across sequential time points (Contrast 1: pre-manipulation = -1, post-manipulation = 1, other time points = 0; Contrast 2: post-manipulation = -1, 24 hours later = 1, other time points = 0; Contrast 3: 24 hours later = -1, one week later = 1, other time points = 0). For ease of interpretation, I also established contrast codes for the additional stimulus race (White = -1, Black = 1) and condition (matched control = -1, holistic training = 1) predictors. To avoid overspecification of a single model by including multiple interactions, I decided to specify a separate multilevel model for each time of test contrast and its interaction with stimulus race and condition. This resulted in three models that predicted recognition memory (*d'*) from a three-way interaction between time of test (Contrast 1, 2, or 3), stimulus race, and condition. I entered all predictors as fixed effects and included participants as a random intercept. Due to contrast coding, coefficients in the models reflect mean differences between the reference group and comparison group for each predictor.

The first multilevel model (Table 3) included the first time of test contrast, which compared pre-manipulation recognition to immediate post-manipulation recognition.

Table 3

Experiment 2 Model of Recognition Memory Strength from Time of Test (Pre- versus Post-Manipulation), Stimulus Race, and Condition

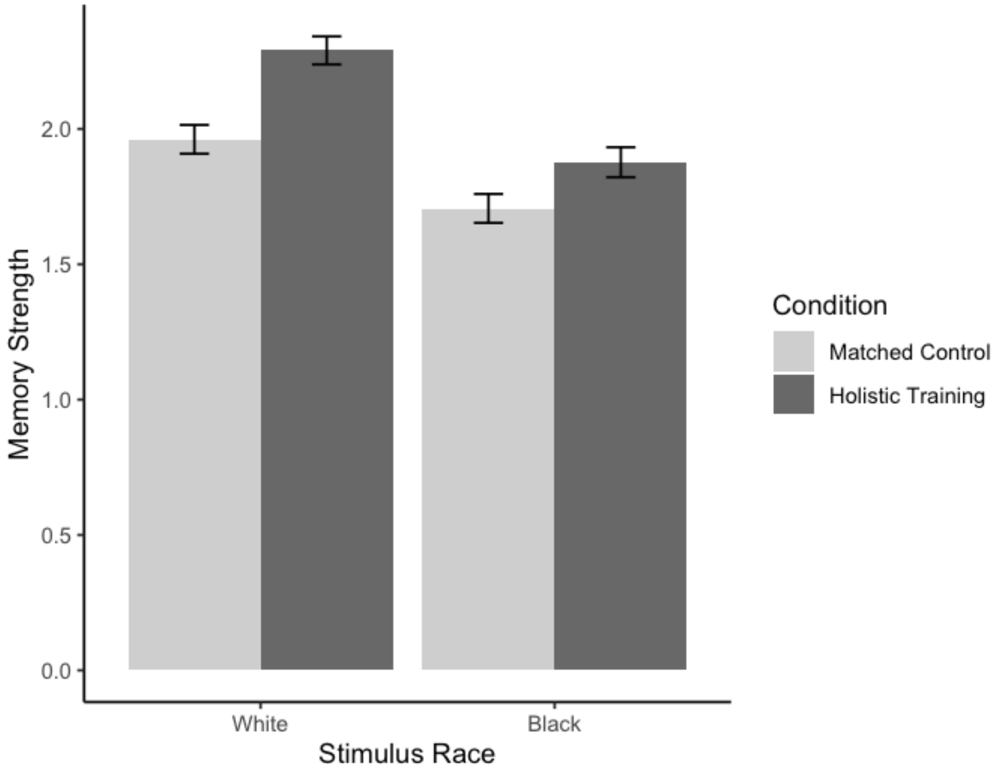
	Estimate	Standard Error
(Intercept)	1.959***	0.055
Time of Test	0.048	0.028
Stimulus Race	-0.167***	0.020
Condition	0.125*	0.055
Time of Test × Stimulus Race	-0.052	0.028
Time of Test × Condition	-0.052	0.028
Stimulus Race × Condition	-0.039*	0.020
Time of Test × Stimulus Race x Condition	-0.008	0.028
AIC	4125.4	
BIC	4179.0	
ICC	0.5	

Note. *$p < .05$, ***$p < .001$. Time of test reflects pre-manipulation versus post-manipulation, such that positive values represent better average recognition after the manipulation. Stimulus race reflects White versus Black faces, such that positive values represent better average recognition of Black faces. Condition reflects matched control versus holistic training, such that positive values represent better average recognition within the holistic training condition.

Main effects of stimulus race and condition were qualified by an interaction between the two variables, such that the difference in recognition of Black and White faces varied by condition, $t(1366) = 2.00$, $p = .05$, 95% CI = [-0.08, -0.0009] (Figure 11). Pairwise comparisons of estimated marginal means indicated that training participants ($M = 2.29$, $SE = 0.08$) recognized White faces better than control participants ($M = 1.96$, $SE = 0.08$), $t(247) = 2.82$, $p = .005$; there was no difference in recognition of Black faces between training and control participants. No other effects in the first model were reliable.

Figure 11

Experiment 2 Stimulus Race by Experimental Condition Interaction



Note. Average recognition memory strength for Black and White faces, separated by condition. Error bars represent standard errors.

The second model included the second time of test contrast, which compared immediate post-manipulation recognition to recognition 24 hours later (Table 4).

Table 4

Experiment 2 Model of Recognition Memory Strength from Time of Test (Post-Manipulation versus 24 Hours Later), Stimulus Race, and Condition

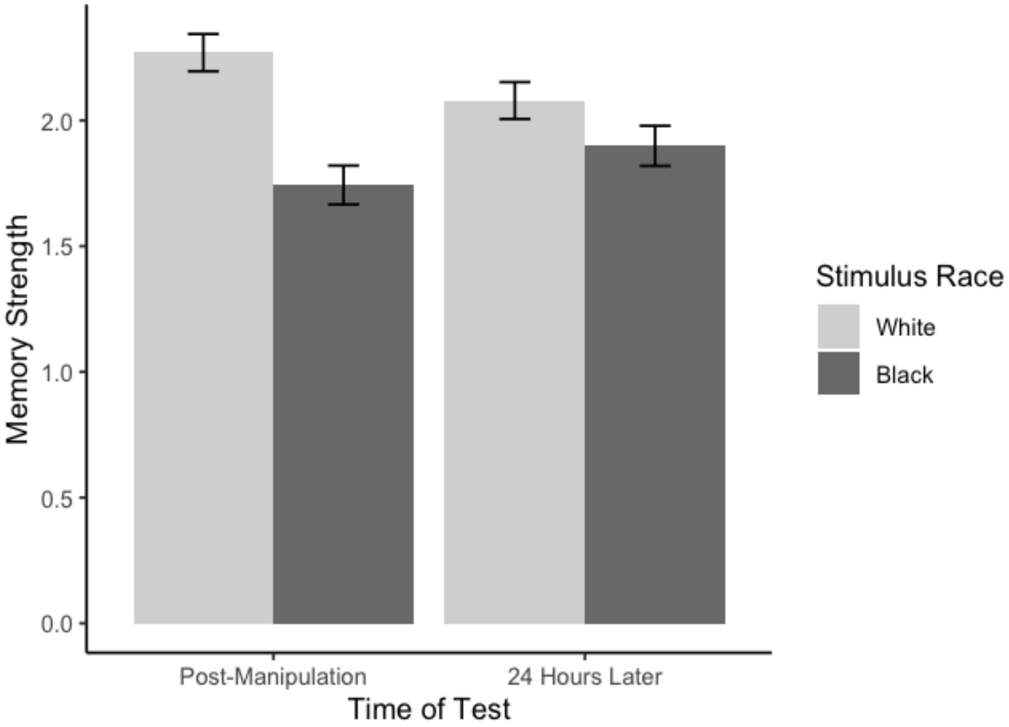
	Estimate	Standard Error
(Intercept)	1.959***	0.055
Time of Test	-0.008	0.028
Stimulus Race	-0.167***	0.020
Condition	0.125*	0.055
Time of Test × Stimulus Race	0.086**	0.028
Time of Test × Condition	0.010	0.028
Stimulus Race × Condition	-0.039*	0.020
Time of Test × Stimulus Race x Condition	-0.004	0.028
AIC	4125.7	
BIC	4179.3	
ICC	0.5	

Note. **p* < .05, ***p* < .01, ****p* < .001. Time of test reflects post-manipulation versus 24 hours later, such that positive values represent better recognition 24 hours later. Stimulus race reflects White versus Black faces, such that positive values represent better recognition of Black faces. Condition reflects matched control versus holistic training, such that positive values represent better recognition within the holistic training condition.

Other than the main effects and interaction between stimulus race and condition seen in the first model, there was an additional interaction between stimulus race and time of test, $t(1366) = 3.10$, $p = .02$, 95% CI = [0.03, 0.14]. Pairwise comparisons of estimated marginal means indicated that participants recognized White faces ($M = 2.22$, $SE = 0.07$) substantially better than Black faces ($M = 1.71$, $SE = 0.07$) immediately after the manipulation, $t(1366) = 7.42$, $p < .001$. However, 24 hours later, better recognition of White faces ($M = 2.03$, $SE = 0.07$) than Black faces ($M = 1.87$, $SE = 0.07$) became considerably less pronounced, $t(1366) = 2.36$, $p = .02$ (Figure 12). No other effects in the second model were reliable.

Figure 12

Experiment 2 Stimulus Race by Time of Test Interaction



Note. Average recognition memory strength for Black and White faces, after the manipulation compared to 24 hours later. Error bars represent standard errors.

The final model included the third time of test contrast, which compared recognition 24 hours later to recognition 1 week later (Table 5). Again, main effects of stimulus race and condition were qualified by an interaction between the two variables. No other effects in the third model were reliable.

Table 5

Experiment 2 Model of Recognition Memory Strength from Time of Test (24 Hours Later versus 1 Week Later), Stimulus Race, and Condition

	Estimate	Standard Error
(Intercept)	1.959***	0.055
Time of Test	-0.036	0.028
Stimulus Race	-0.167***	0.020
Condition	0.125*	0.055
Time of Test × Stimulus Race	-0.028	0.028
Time of Test × Condition	0.039	0.028
Stimulus Race × Condition	-0.039*	0.020
Time of Test × Stimulus Race x Condition	0.025	0.028
AIC	4129.9	
BIC	4183.5	
ICC	0.5	

Note. **p* < .05, ****p* < .001. Time of test reflects 24 hours later versus 1 week later, such that positive values represent better recognition 1 week later. Stimulus race reflects White versus Black faces, such that positive values represent better recognition of Black faces. Condition reflects matched control versus holistic training, such that positive values represent better recognition within the holistic training condition.

Memory Self-Efficacy and Recognition Memory

Participants responded to the memory self-efficacy items after the manipulation (*M* = 5.69, *SD* = 2.22), 24 hours later (*M* = 5.75, *SD* = 2.02), and 1 week later (*M* = 5.78, *SD* = 2.03). As an initial test of my predictions regarding the connection between memory self-efficacy and face recognition, I first correlated average memory self-efficacy scores with average White *d*', and separately, average Black *d*', at each testing occasion. Memory self-efficacy was positively related to recognition of White faces at all time points, *r'*s = 0.15–0.20, *p*'s = .004–.02. Memory self-efficacy was also positively related to recognition of Black faces at all time points, but the correlations were not reliable, *r'*s = 0.05–0.13, *p*'s = .07–.47.

To explore whether recognition was potentially impacted by a joint influence of memory self-efficacy and another predictor(s), I repeated the latter two multilevel models above and included an interaction term of memory self-efficacy (centered within-cluster), along with memory self-efficacy as a random slope. I did not feel it was relevant to repeat the first model (i.e., comparing pre-manipulation to post-manipulation) because memory self-efficacy was the focus of these additional analyses, and I did not measure memory self-efficacy prior to the manipulation.

The first multilevel model included the first time of test contrast, which compared recognition immediately after the manipulation and 24 hours later. The model revealed a main effect of memory self-efficacy, such that greater belief in one's face recognition abilities corresponded with better recognition, *t*(72.36) = 2.46, *p* = .02, 95% CI = [0.02, 0.22]. As in earlier models, there was also a main effect of stimulus race and interactions between stimulus race and condition, and stimulus race and time of test. No other effects were reliable.

The second multilevel model included the second time of test contrast, which compared recognition 24 hours later to 1 week later. The model revealed a main effect of memory self-efficacy once again, such that greater belief in one's face recognition abilities corresponded with better recognition. There was also a main effect of stimulus race again. No other effects were reliable.

Discussion

My hypotheses for Experiment 2 were partially supported. Though training participants displayed better recognition overall than control participants, the baseline difference in recognition between the two conditions suggests that this main effect of condition was due to a randomization failure; specifically, it appears that training participants were initially better at recognizing faces than control participants, rather than the training boosting recognition among these participants. It is unclear why a randomization failure occurred, given that all participants had an equal opportunity of being assigned to either condition. However, this explanation is further supported by the fact that recognition did not reliably change over time among training participants; if the training was effective, recognition would have improved at subsequent time points relative to baseline, which the data did not show. Even more puzzling is that control participants demonstrated reliable improvements in recognition over time while training participants did not. Though these improvements ultimately returned to baseline by the final measurement occasion, it is still strange that this occurred within the control condition rather than the training condition, as would be expected.

The pattern of results in this experiment differ considerably in comparison to Experiment 1. Both conditions in Experiment 2 displayed improved recognition for White faces immediately after the manipulation (and reliably so in the control condition), yet both conditions in Experiment 1 displayed almost no difference in recognition for White faces before versus after the manipulation. Experiment 2 training participants showed a descriptive decline in recognition of Black faces before versus after the manipulation, which is similar in direction to both Experiment 1 conditions; and yet, Experiment 2 control participants showed a descriptive *increase* in recognition of Black faces before versus immediately after the manipulation. These inconsistencies suggest that the training is not refined enough to result in predictable recognition outcomes.

In a departure from Experiment 1, I recruited White participants for this experiment to begin exploring the impact of my training on other-race (i.e., Black, in this case) face recognition. Though there were descriptive improvements in Black recognition among both conditions from immediately after the manipulation to 24 hours later, recognition of Black faces had returned to baseline in both conditions 1 week later. This suggests that, though the intervention may have short-term effects on recognition, the effects are not sustained.

One outcome that did align with my initial hypothesis was the positive relationship between memory self-efficacy and face recognition. It is important to note, however, that memory self-efficacy was measured *after* the recognition memory tasks, so I cannot claim at this point that higher memory self-efficacy leads to better face recognition. Rather, this result may – at least in part – reflect participants' judgments of their performance on the memory task they just completed. If interpreted in this way, the stronger relationships between memory-self efficacy and same-race recognition (in comparison to the relationship with other-race recognition) could suggest that individuals may base their memory self-efficacy judgments more so on their perceived ability to recognize same-race, rather than other-race faces. This is merely speculation, however, since my assessment of memory self-efficacy was not race-specific.

In sum, Experiment 2 offers additional evidence that the training does not improve recognition above and beyond the matched control. As I alluded to in the Experiment 1 Discussion, it appears that the conditions are not distinct enough from one another to result in markedly different outcomes. Further, it appears that neither condition produces lasting change in face recognition beyond 24 hours; this is understandable, given that the intervention only requires approximately 30-45 minutes and is completed in a single testing session. It would be interesting to explore whether a more prolonged intervention, consisting of several training sessions spanning multiple days, would have a more lasting impact.

Experiment 3

In developing my face recognition training, one issue I hoped to improve is poor recognition for other-race faces. People generally experience difficulties distinguishing between faces of another race, yet tend to have greater ease doing so for same-race faces; this common experience is called the cross-race effect (CRE), and can lead to an assortment of adverse real-world outcomes (e.g., eyewitness misidentification - Wells & Olson, 2001; Wilson et al., 2013; social discomfort - McKone et al., 2023).

In Experiment 3, I sought to examine the effectiveness of my training in improving other-race face recognition relative to two other established interventions designed for this purpose: subordinate-level individuation (Tanaka & Pierce, 2009) and unguided CRE awareness (Hugenberg et al., 2007). Though these approaches and others that currently exist in the CRE literature provide theoretical demonstrations that other-race face recognition can be improved, none to my knowledge can be readily applied in real-world settings to enact practical change. If my training was comparable to, or better than, two prominent interventions in improving other-race face recognition, it would provide evidence that it is a viable option in mitigating the cross-race effect; and importantly, given that skills learned in the training should be applicable in everyday life, that the training may have the capacity improve face recognition in practical applications.

To explore this question, I recruited Black and White online participants residing in the United States. Because the training materials and memory tasks consist of Black and White faces, constraining the Experiment 3 sample to Black or White individuals would shed light on how the training influences recognition memory for same- versus other-race faces.

I anticipated that White participants would display better memory for White than Black faces when measured before the manipulation; for Black participants however, both possibilities were plausible. Black participants have displayed the expected ingroup recognition advantage for Black faces on some occasions (Dodson & Dobolyi, 2015; Pauker et al., 2009), yet on other occasions they have better recognized White than Black faces (Pica et al., 2015; Simon et al., 2023; Wright et al., 2003).

Regardless of which race Black participants better recognized prior to the manipulation, I expected that both Black and White participants would display improved memory for Black faces after the manipulation. Though this improvement is plausible for all conditions, I anticipated the most pronounced improvement would be among participants in the holistic training condition compared to participants in the other two training conditions.

After the manipulation, holistic training participants could display improvement in White face recognition as well. In that case, it would be interesting to note how the difference between same- and other-race face recognition changed after the manipulation: recognition could improve for both races in similar amounts, and the size of the gap would remain the same; alternatively, recognition for one group could occur to a greater extent than the other, altering the size of the gap.

Method

Study Design

This experiment was a 2x2x2x3 mixed design, including within-subjects factors of stimulus race (Black, White) and time of testing (before manipulation, after manipulation), and between-subjects factors of participant race (Black, White) and training condition (holistic training, subordinate-level individuation training, unguided CRE awareness training).

Participants

I recruited non-Latinx, White and Black adults residing in the United States through CloudResearch Connect (*N* = 274). As compensation, participants received $15 for an expected 1 hour of participation. An a priori power analysis indicated that, with 48 observations per person, I would have 80% power to capture an effect size of at least .14 with *N* = 112 participants. Considering that this experiment would contain three experimental conditions rather than two, I aimed to recruit at least 2.25x this amount (i.e., *N* = 252 participants).

Removing individuals who did not pass the quality control or attention checks (*n* = 27), had incomplete data (*n* = 5), or did not identify as exclusively Black or White (*n* = 9) resulted in a final sample of 255 participants (M_{age} = 37.63, *SD* = 11.67). There was an approximately even split of Black (*n* = 131) relative to White (*n* = 124) participants in this final sample. 79 of the participants completed the holistic training, 89 completed the individuation training, and 87 completed the CRE awareness training.

Materials and Manipulation

Face Stimuli. I used the Experiment 1 Chicago Face Database stimuli for the experimental conditions and pre- and post-manipulation recognition memory tasks. Though the subordinate-level individuation condition required fewer unique faces than the holistic training and CRE awareness conditions, all faces depicted in the individuation condition were seen by participants in the other two conditions as well.

Experimental Manipulation. Participants completed one of three different conditions: holistic training, subordinate-level individuation, or CRE awareness. The holistic training condition was identical to those in Experiments 1 and 2.

The second condition, subordinate-level individuation[1], was adapted from Tanaka and Pierce (2009). In alternating learning and naming blocks, participants learned to associate 8 Black and 8 White male faces with letters on the keyboard, then were quizzed on their memory of these associations. Other-race faces, based on participants' racial identity indicated in the initial demographics questionnaire, were each paired with a unique letter (e.g., "A","S", "D"...). Alternatively, same-race faces were all paired with the same letter, "O". The first learning block presented two White and two Black faces, and subsequent learning blocks introduced one new face of each race until a total of 8 faces per race were learned. In the naming blocks that followed each learning block, the faces that had been introduced up to that point were presented one at a time without their labels. Participants received feedback after each trial specifying whether or not they answered correctly; if they were incorrect, they were also shown the correct label. Participants were required to enter the correct letters for all faces in the block to progress, or the block would repeat up to five times. The individuation condition concluded when participants successfully completed all learning and naming blocks.

The third condition, unguided CRE awareness, was modeled after Hugenberg and colleagues (2007). This intervention began by informing participants about the existence of the CRE with the following instructions:

---

[1] The subordinate-level individuation training developed by Tanaka and Pierce (2009) contained an additional timed-response task after the learning and naming blocks. In an effort to keep the three training conditions similar in length, I omitted the timed-response task portion of the training for Experiment 3.

> Previous research has shown that people reliably show what is known as the Cross-Race Effect (CRE) when learning faces. Basically, people tend to confuse faces that belong to other races. For example, a White learner will tend to mistake one Black face for another. Now that you know this, we would like you to try especially hard when learning faces in this task that happen to be of a different race. Do your best to try to pay close attention to what differentiates one particular face from another face of the same race, especially when that face is not of the same race as you.
>
> Remember, pay very close attention to the faces, especially when they are of a different race than you in order to try to avoid this Cross-Race Effect.

After these instructions, participants completed a lineup recognition paradigm consisting of a study phase, distractor task, and test phase (Meissner et al., 2005)[2]. In the study phase, participants viewed 7 Black and 7 White male faces one at a time for three seconds each. After completing a five-minute distractor task, participants viewed a total of 28 lineups, 14 for each race. Each lineup depicted 3 different faces, with half of lineups containing a target face and half of them not. For each lineup, participants either selected a face they believed they saw previously, indicated that a target face was not present, or stated that they were unsure. They also rated their confidence in their identification decision for each lineup.

Procedure

After providing informed consent, participants began the experiment by completing two quality control checks; only individuals who correctly responded to these checks were able to continue. Participants then completed the demographics

---

[2] In Hugenberg et al. (2007), participants completed a recognition memory task after receiving CRE awareness instructions. Since all participants would complete a post-manipulation recognition memory task in Experiment 3, I included a lineup recognition paradigm after the CRE awareness instructions instead.

questionnaire before beginning the main portion of the study. The remainder of the testing session was largely identical to Experiment 1: participants completed a baseline recognition memory task, followed by their randomly-assigned experimental condition, then concluded with a second recognition memory task and debriefing.

Results

There were three levels of nesting within the data: stimuli (two Level 1 units, Black and White) within time points (two Level 2 units, pre- and post-manipulation) within participants (254 Level 3 clusters). Each Level 3 cluster (i.e., participant) contained the same number of Level 1 and Level 2 units.

Change in Recognition Memory by Condition

From trial-level “yes, no” responses on the two recognition memory tests, I tallied participants’ number of hits, false alarms, misses, and correction rejections for Black and White faces. I then computed each participant’s hit and false alarm rate (i.e., number of hits/false alarms divided by the number of possible hits/false alarms) for the two memory tests, and adjusted for rates of 0 or 1 by adding or subtracting 0.5. Within the two tests, hit rates of 0 occurred twice and 3 times, false alarm rates of 0 occurred 164 and 189 times, hit rates of 1 occurred 125 and 110 times, and false alarm rates of 1 occurred once and twice. Descriptive statistics for hits and false alarm rates can be found in Appendix D.

Finally, I computed each participant’s average memory strength (operationalized as signal detection parameter *d*’) for Black and White faces during the tests. The number of participants with *d*’ below zero across trials, indicating chance performance throughout the experiment, can be found in Appendix A. Descriptive statistics of White and Black recognition for all conditions before and after the manipulation can be found in Table 6.

Table 6

Experiment 3 Descriptive Statistics

	Before Manipulation				**After Manipulation**			
	Black Faces		**White Faces**		**Black Faces**		**White Faces**	
Condition	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Holistic Training	2.05^{a}	1.15	2.17	0.92	$1.81^{a,j}$	0.98	1.97	0.99
Black Participants	2.09	1.19	1.97	0.94	1.95	0.84	1.86	0.96
White Participants	$2.02^{b,d}$	1.11	2.36^{b}	0.87	$1.67^{c,d}$	1.09	2.07^{c}	1.02
Individuation	1.86^{e}	1.20	2.09^{e}	1.09	2.08	1.20	2.01	1.25
Black Participants	1.99	1.20	2.10	1.18	2.25	1.13	2.11	1.28
White Participants	1.67^{f}	1.19	2.08^{f}	0.94	1.82	1.26	1.86^{k}	1.20
CRE Awareness	1.97^{g}	1.09	2.20^{g}	1.01	2.16^{j}	1.16	2.27	1.14
Black Participants	2.20	1.01	2.08	1.12	2.31	1.16	2.11	1.24
White Participants	1.79^{h}	1.13	2.30^{h}	0.91	2.05^{i}	1.16	$2.40^{i,k}$	1.05

Note. Means with shared superscripts are reliably different from one another. Shared superscripts in the same row indicate a reliable difference between stimulus races or time points within the same condition. Shared superscripts in the same column indicate a reliable difference between conditions for a particular stimulus race within a single time point.

A one-way ANOVA indicated that participants in the three conditions did not meaningfully differ in their baseline recognition of White faces, $F(2, 251) = 0.28$, $p = .76$, nor in their recognition of Black faces, $F(2, 251) = 0.56$, $p = .57$. All conditions better recognized White than Black faces prior to the manipulation, but the difference was only reliable among individuation condition participants, $t(88) = 1.97$, $p = .05$, $d = .20$, and CRE awareness participants, $t(86) = 2.21$, $p = .03$, $d = 0.22$. Better recognition of White than Black faces before the manipulation was not reliable among holistic training participants, $t(78) = 0.97$, $p = .34$, $d = .11$.

Prior to the manipulation, Black participants in the holistic training ($n = 39$) and CRE awareness ($n = 38$) conditions displayed better recognition of Black than White faces (i.e., a descriptive cross-race effect); alternatively, Black participants in the individuation condition ($n = 54$) displayed the reverse, descriptively better recognizing White than Black faces prior to the manipulation. White participants, on the other hand, recognized White faces reliably better than Black faces in all three conditions (holistic $n = 40$; individuation $n = 35$; CRE awareness $n = 49$), ts < 3.89, ps < .05, ds < 0.48.

After the manipulation, individuation participants displayed descriptively better recognition of Black faces relative to their Black recognition at baseline, $t(88) = 1.62$, $p = .11$, $d = 0.18$. CRE awareness participants also displayed a descriptive improvement in Black recognition after the manipulation, $t(86) = 1.57$, $p = .12$, $d = 0.17$. Alternatively, training participants displayed reliably *worse* recognition of Black faces relative to their Black recognition at baseline, *t*(78) = 2.03, *p* = .05, *d* = 0.23.

Accounting for participants' race in post-manipulation recognition, Black participants in all three conditions displayed a descriptive cross-race effect, better recognizing Black than White faces after the manipulation. White participants in the holistic training and CRE awareness conditions continued to recognize White faces reliably better than Black faces after the manipulation, *t*s < 3.00, *p*s < .05, *d*s < 0.37. In contrast, White individuation participants displayed almost no difference in recognition of White and Black faces, *t*(34) = 0.26, *p* = .79, *d* = 0.04.

Primary Analyses

The intraclass correlation (ICC) of the unconditional model was .49, indicating that there was sufficient variability at Level 3 to justify the use of multilevel modeling. I employed contrast coding once again to specify different control conditions as the reference group against my holistic training. I used two contrasts to compare participants' recognition memory based on their condition (Contrast 1: individuation = -1, holistic training = 1, CRE awareness = 0; Contrast 2: CRE awareness = -1, holistic training = 1, individuation = 0). For ease of interpretation, I also established contrast codes for the additional stimulus race (White = -1, Black = 1), time of test (pre-manipulation = -1, post-manipulation = 1), and participant race White = -1, Black = 1) predictors.

As in Experiment 2, I specified a separate multilevel model for each condition contrast and its interaction with stimulus race, time of test, and participant race. This resulted in two models that predicted recognition memory (*d'*) from a four-way interaction between condition (Contrast 1 or 2), stimulus race, time of test, and participant race. I entered all predictors as fixed effects and included participants as a random intercept. Due to contrast coding, coefficients in the models reflect mean differences between the reference group and comparison group for each predictor.

The models were fit using restricted maximum likelihood estimation with the *lme4* package (Bates et al., 2015) in R version 4.3.2 (R Core Team, 2023). Inferential tests of the fixed effects used the Satterthwaite method of degrees of freedom approximation, computed with the *lmerTest* package (Kuznetsova et al., 2017).

The first model (Table 7) compared the holistic training to the individuation condition.

Table 7

Experiment 3 Model of Recognition Memory Strength from Time of Test, Stimulus Race, Participant Race, and Condition (Individuation versus Holistic Training)

	Estimate	Standard Error
(Intercept)	2.050***	0.055
Time of Test	-0.002	0.024
Stimulus Race	-0.065**	0.024
Participant Race	0.031	0.055
Condition	0.008	0.068
Time of Test × Stimulus Race	0.033	0.024
Time of Text x Participant Race	0.015	0.024
Stimulus Race x Participant Race	0.110***	0.024
Time of Test × Condition	-0.070*	0.030
Stimulus Race × Condition	-0.005	0.030
Participant Race x Condition	-0.074	0.068
Time of Test × Stimulus Race x Participant Race	-0.006	0.024
Time of Test × Stimulus Race x Condition	-0.043	0.030
Time of Test × Participant Race x Condition	0.006	0.030
Stimulus Race × Participant Race x Condition	0.029	0.030
Time of Test × Stimulus Race x Participant Race x Condition	0.008	0.030
AIC	2890.4	
BIC	2979.1	
ICC	0.5	

Note. **p* < .05, ***p* <.01, ****p* < .001. Time of test reflects pre-manipulation versus post-manipulation, such that positive values represent better recognition after the manipulation. Stimulus race reflects White versus Black faces, such that positive values represent better recognition of Black faces. Participant race reflects White versus Black participants, such that positive values represent better recognition among Black participants. Condition reflects individuation versus holistic training, such that positive values represent better recognition within the holistic training condition.

The model revealed a main effect of stimulus race that was qualified by an interaction with participant race, such that the difference in recognition for Black and White faces varied depending on participants' race (Figure 13), $t(753) = 4.49$, $p < .001$, 95% CI = [0.06, 0.16]. Pairwise comparisons of estimated marginal means indicated that White participants, on average, recognized White faces ($M = 2.19$, $SE = 0.09$) better than Black faces ($M = 1.84$, $SE = 0.09$), $t(750) = 4.79$, $p < .001$; alternatively, Black participants did not differ in their average recognition of White versus Black faces.

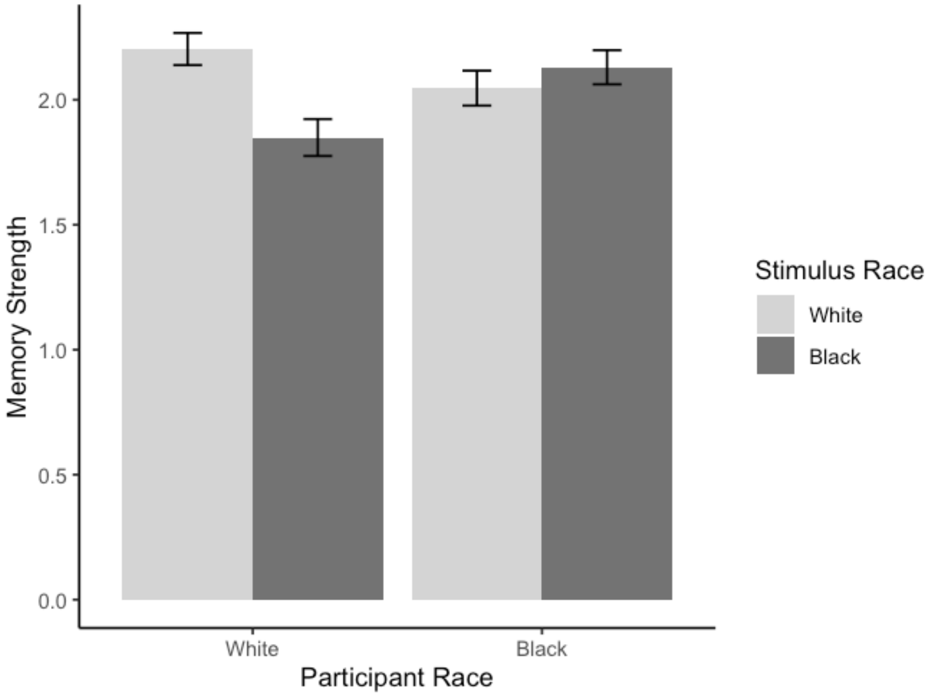


Figure 13

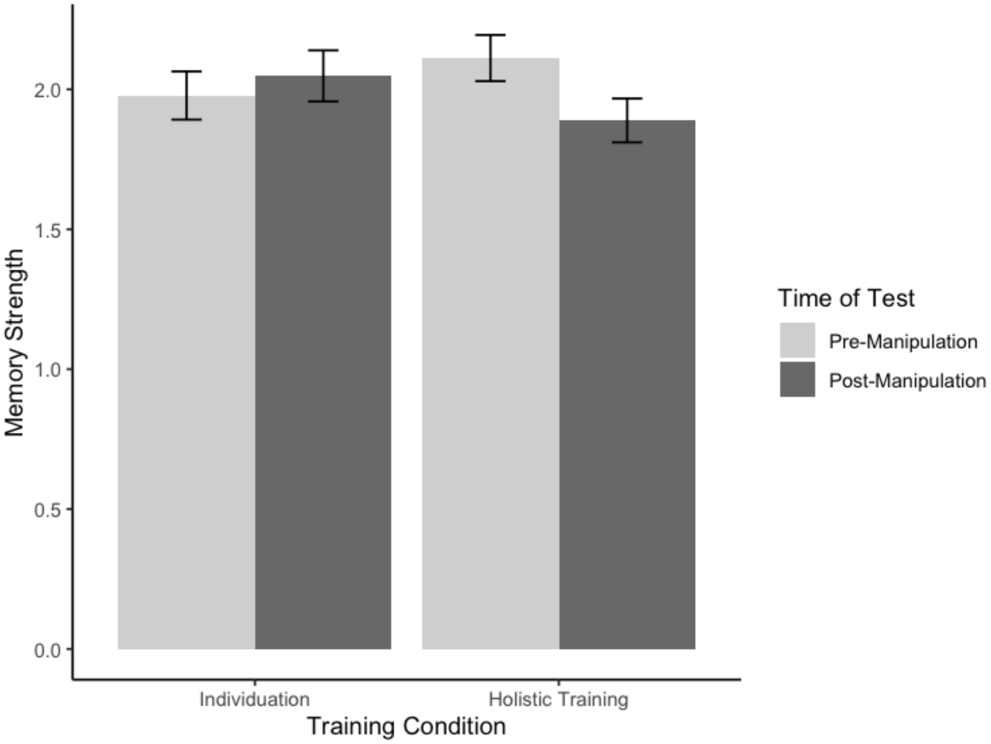
Experiment 3 Participant Race by Stimulus Race Interaction

Note. Average recognition memory strength among Black and White participants, separated by stimulus race. Error bars represent standard errors.

There was also an interaction between time point and condition, such that the difference in recognition before versus after the manipulation varied depending on participants' experimental condition (Figure 14), $t(753) = 2.32$, $p = .02$, 95% CI = [-0.13, -0.01]. Pairwise comparisons of estimated marginal means indicated that training participants' average recognition after the manipulation was worse ($M = 1.99$, $SE = 0.10$), on average, than their recognition before the manipulation ($M = 2.13$, $SE = 0.10$), $t(753) = -1.83$, $p = .07$; alternatively, individuation participants displayed the opposite: their average recognition after the manipulation was better, on average, than their recognition before, $t(753) = 1.77$, $p = .08$. However, neither of these differences were reliable.

Figure 14

Experiment 3 Condition (Individuation versus Holistic Training) by Time of Test Interaction



Note. Average recognition memory strength among individuation and holistic training participants, before and after the manipulation. Error bars represent standard errors.

The second model compared the holistic training to the CRE awareness condition (Table 8).

Table 8

Experiment 3 Model of Recognition Memory Strength from Time of Test, Stimulus Race, Participant Race, and Condition (CRE Awareness versus Holistic Training)

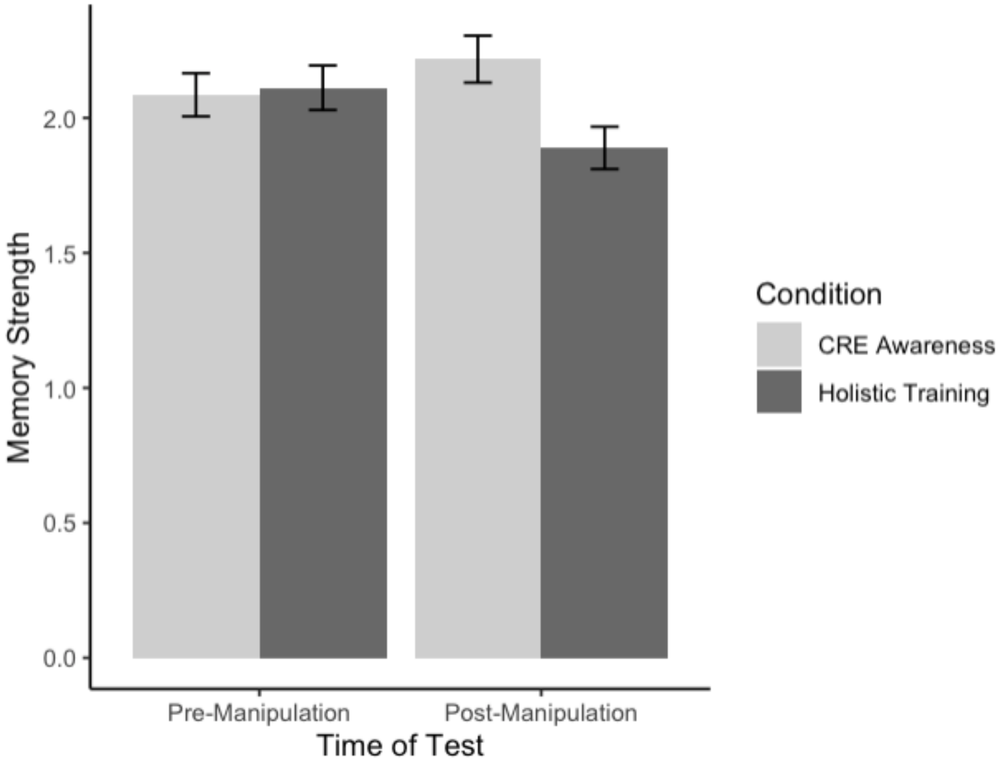
	Estimate	Standard Error
(Intercept)	2.054***	0.055
Time of Test	-0.005	0.024
Stimulus Race	-0.067**	0.024
Participant Race	0.035	0.055
Condition	-0.080	0.068
Time of Test × Stimulus Race	0.033	0.024
Time of Text x Participant Race	0.024	0.024
Stimulus Race x Participant Race	0.108***	0.024
Time of Test × Condition	-0.087**	0.030
Stimulus Race × Condition	0.002	0.030
Participant Race x Condition	-0.022	0.068
Time of Test × Stimulus Race x Participant Race	-0.002	0.024
Time of Test × Stimulus Race x Condition	-0.019	0.030
Time of Test × Participant Race x Condition	0.038	0.030
Stimulus Race × Participant Race x Condition	-0.015	0.030
Time of Test × Stimulus Race x Participant Race x Condition	0.005	0.030
AIC	2890.4	
BIC	2979.1	
ICC	0.5	

Note. **p* < .05, ***p* <.01, ****p* < .001. Time of test reflects pre-manipulation versus post-manipulation, such that positive values represent better recognition after the manipulation. Stimulus race reflects White versus Black faces, such that positive values represent better recognition of Black faces. Participant race reflects White versus Black participants, such that positive values represent better recognition among Black participants. Condition reflects CRE awareness versus holistic training, such that positive values represent better recognition within the holistic training condition.

Like in the first model, a main effect of stimulus race was qualified by an interaction with participant race. There was also an interaction between time of test and condition (Figure 15); *t*(753) = -2.87, *p* = .004, 95% CI = [-0.15, -0.03]. Pairwise comparisons of estimated marginal means indicated that participants' recognition in the two conditions did not differ before the manipulation; however, recognition was substantially worse among holistic training participants (*M* = 1.88, *SE* = 0.10) than CRE awareness participants (*M* = 2.22, *SE* = 0.09) after the manipulation, *t*(355) = 2.24, *p* = .03.

Figure 15

Experiment 3 Time of Test by Condition (CRE Awareness versus Holistic Training) Interaction



Note. Average recognition memory strength before and after the manipulation, separated by condition (CRE awareness versus holistic training). Error bars represent standard errors.

Discussion

My intent with Experiment 3 was to compare the holistic training against other established interventions in their ability to influence other-race face recognition. As is typically demonstrated in cross-race effect literature, White participants better recognized same- versus other-race faces across conditions and measurement occasions, whereas recognition was relatively mixed among Black participants. In past work, I have proposed that non-White people may be motivated to attend to White faces due to their practical importance (Simon et al., 2023); social and material power are largely held by White individuals in the United States, and people pay attention to those who are relevant to their outcomes (Fiske, 1993; Neuberg & Fiske, 1987). Thus, when Black Americans view Black and White faces, the motivational pull of the ingroup conflicts with the motivational pull of the higher-status outgroup, which in turn produces competing attentional responses. However, when White Americans view Black and White faces, the two motivations complement one another and work in unison to direct attention toward White faces. General findings from Experiment 3 provide further support for this theory.

My primary hypothesis for this experiment was that other-race recognition would improve after the manipulation for holistic training participants. This hypothesis was not supported; instead, training participants' other-race recognition worsened after the manipulation, whereas participants in the other conditions displayed improved other-race recognition relative to baseline. This outcome coincides with results from Experiments 1 and 2; rather than improving recognition, my holistic training appears to produce the opposite effect.

General Discussion

My ultimate objective when developing this face recognition training was to offer a practical tool people could use to better remember faces in their daily lives. I based the training on the most variable facial features across ethnic groups, so that the techniques – if effective – could be applied to a wide range of faces. I focused on improving holistic processing, a form of visual processing known to correspond with accurate face recognition. To maximize the likelihood of its efficacy, I integrated two evidence-based instructional approaches employed by educators to boost student learning. Despite the consideration that went into developing this training, the three experiments I conducted appear to convey that the training, in its current form, does not improve face recognition.

In Experiment 1, training participants' recognition of White faces did not change before versus after the manipulation, whereas their recognition of Black faces declined after the training. In Experiment 2, White training participants demonstrated slight fluctuations in same- and other-race recognition over time, but ultimately returned to their baseline recognition after one week. In Experiment 3, two previously established interventions incited better recognition of faces after the manipulation, whereas my training led to poorer recognition.

The recognition declines participants displayed immediately after training in Experiments 1 and 3, though initially surprising, replicate findings from some early interventions (Malpass, 1981; Woodhead et al., 1979); interestingly, these interventions also tasked participants with attending to facial features. Though my training differed from their approaches by focusing on the spatial relationships between features, the consistency in findings further backs my suspicion that the training did not increase holistic processing.

Some have suggested that participants may experience difficulty attempting to implement a new encoding approach because they've processed faces a certain way their whole lives, and old habits are hard to break, resulting in poorer performance on the post-intervention recognition task (Malpass, 1981). A few examples of qualitative feedback from training participants in my studies speak to this explanation; for example, one participant wrote, “I couldn’t really implement the techniques you taught us in the first lesson… In the first session, I gave everyone a nickname based upon a feature or if they looked like a celebrity or a person I know and I was more confident using that strategy.” By assessing participants at multiple time points in Experiment 2, I attempted to gauge whether recognition would improve over time if participants continued implementing the training techniques outside the experiment; however, as alluded to by this individual, I had no control over whether participants applied the training material to subsequent recognition. Therefore, though it may have felt difficult and unnatural to some participants initially, it is unclear whether recognition would improve if individuals remained persistent in trying to adopt the techniques.

The recognition memory task employed in this research is a lab-based paradigm that has long been critiqued for conflating picture recognition with face recognition (Bruce & Young, 1986); in other words, it does not account for external factors such as viewpoint, facial expression, and lighting that influence our recognition of faces in the real world. Indeed, participants fail to recognize previously-viewed, unfamiliar faces if they are presented at a slightly different angle or with a different expression during the recognition task (Bruce, 1982; Longmore et al., 2008; Newell et al., 1999); and yet, this paradigm remains one of the most popular outcome measures in face recognition research (Singh et al., 2022). Though my training did not improve face recognition as defined by this task, the intent behind the training is to improve face recognition in applied contexts; therefore, perhaps a more realistic assessment would be better suited to assess its efficacy. Newer face recognition tasks have begun to account for external factors that influence face recognition in real-world settings (e.g., the Cambridge Face Memory Test presents faces from different angles - Duchaine & Nakayama, 2006), so employing one of these tasks may be worthwhile to better assess the efficacy of the training.

Nevertheless, the question still remains as to why the individuation and CRE awareness interventions in Experiment 3 improved recognition (albeit of pictures) and my training did not. Upon examining the average time each condition took to complete the study, I discovered that holistic training participants took substantially longer to complete the study on average than the other two conditions. Further, the other conditions did not attempt to teach participants a new approach to encoding faces. As I mentioned above, I suspect that the training incited fatigue and contained material that was difficult for participants to adopt, which led to poorer recognition on the memory task relative to the other conditions.

Another point of intrigue is the discrepancy between recognition outcomes and the majority of participants' qualitative feedback about the training. Although training participants often did not display improvement in face recognition, many left comments stating that they found the intervention meaningful and informative. Even on the last measurement occasion one week after the training, Experiment 2 participants left comments such as, "Enjoyed doing this - still remember and have tried to use the various face features given to me last week" and "I do feel like I'm better at remembering the different faces now compared to when I started the training part of the first study, I wonder if my score actually improved or not." It could be possible that some participants saw the value of the training, yet in actuality were unable to employ the techniques appropriately. Many participants also expressed a desire to know their accuracy scores, which I withheld in order to reduce temptation to use another technique if they weren't performing well; if the training is able to produce improved recognition after some refining though, I would be open to sharing participants' accuracy scores. Doing so would relax the degree of experimental control I would have in the study, but I believe participants' awareness of their recognition ability throughout the experiment would make the training and overall experience more impactful.

When beginning this research, I predicted that training participants may not display improved recognition for a White faces due to a ceiling effect; however, I did not anticipate that training participants would display what appeared to be a ceiling effect for all faces. Even when recognition somewhat improved in Experiment 2, it returned to baseline levels by the end of the experiment. Though there has been a lack of success among face recognition interventions directed toward the general population, interventions for populations with recognition impairment or deficits – including those that target other-race face recognition – have been more promising. Despite the lack of recognition improvement within my own experiments, I am not ready to adopt the perspective that face recognition is a fixed skill that cannot be modified. However, there were limitations to this research and my training that will need to be addressed.

Limitations

One limitation of my training is that it may have incited fatigue in participants. Though each lesson highlighted a different FINNM feature, repeating the same activities within the lessons may have led to decreased attention and focus among participants. In the future, spreading the training across multiple days may be an effective way of keeping participants engaged and pace their learning; participants may also feel more capable of integrating the concepts into their regular face perception if they learned them gradually.

Another limitation pertained to the matched control condition in Experiments 1 and 2. Though I believed that drawing attention to the spatial relationships between facial features would distinguish the holistic training enough from the matched control, the lack of difference in performance between the groups of participants conveyed otherwise. Importantly, my experiments lacked a method of assessing the extent to which participants engaged in holistic processing. I chose the most common recognition memory paradigm as the outcome measure because I was interested in whether my training could improve face recognition; though holistic processing is said to be implicated in face recognition, there are other tasks (e.g., the part-whole task – Tanaka & Farah, 1993; the composite task – Young et al., 1987) that could have more directly shed light on the extent to which participants engaged in holistic processing.

Future Directions

There are several possibilities regarding the future of this research. One of my next endeavors will be to compare and contrast qualitative responses between the two conditions in Experiments 1-2. The quantitative data suggested that there was a lack of distinction between the holistic training and matched control; since both conditions contained free response items, I have the ability to compare how participants qualitatively responded in each condition. If the training condition did facilitate holistic processing, I would expect these participants to use more holistic language in their responses (i.e., identify spatial relationships between features, discuss a face's configuration, etc.). Coding these qualitative responses will be a perfect task for future undergraduates in my lab at Lafayette College.

I would also like to explore memory self-efficacy further, focusing specifically on memory self-efficacy in terms of other-race faces. In Experiment 2, I asked about memory self-efficacy with regard to recognizing faces in general; however, this does not allow me to make claims about how memory self-efficacy may relate to race-specific face recognition. I am interested in exploring whether feelings of incompetence in recognizing other-race faces can negatively impact recognition of that race. Older adults informed that they are about to complete a memory task perform worse on the task than older adults who do not receive these instructions, but only if they possess low memory-self efficacy (Desrichard & Köpetz, 2005); an experiment such as this could easily be replicated in the context of recognition memory. If the same pattern is shown, future work could attempt to improve recognition of faces by improving people's memory self-efficacy.

Finally, if this training does demonstrate an ability to improve face recognition, it will be important to test whether it can improve recognition of faces from other racial/ethnic groups and gender presentations; doing so will confirm its utility for improving recognition of all types of faces. Though my initial test of this training did not pan out as anticipated, with further refining I hope it will prove itself to be beneficial for improving face recognition.

References

Bate, S., Bennetts, R., Mole, J. A., Ainge, J. A., Gregory, N. J., Bobak, A. K., & Bussunt, A. (2015). Rehabilitation of face-processing skills in an adolescent with prosopagnosia: Evaluation of an online perceptual training programme. *Neuropsychological Rehabilitation*, *25*(5), 733–762. https://doi.org/10.1080/09602011.2014.973886

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaudoin, M., & Desrichard, O. (2017). Memory self-efficacy and memory performance in older adults: The mediating role of task persistence. *Swiss Journal of Psychology*, *76*(1), 23–33. https://doi.org/10.1024/1421-0185/a000188

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241. https://doi.org/10.1037/a0022106

Berry, J. M., & West, R. L. (1993). Cognitive self-efficacy in relation to personal mastery and goal setting across the life span. *International Journal of Behavioral Development*, *16*(2), 351–379. https://doi.org/10.1177/016502549301600213

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*, 105-116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

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

Chernorizov, A. M., Jiang, Z., Petrakova, A. V., & Zinchenko, Y. P. (2016). Face cognition in humans: Psychophysiological, developmental, and cross-cultural aspects. *Psychology in Russia: State of the Art*, *9*(4), 37-50.

Cruz, F., Palma, T. A., Bansemer, E., Correll, J., Fonseca, S., Gonçalves, P., & Santos, A. S. (2023). Do individuation instructions reduce the cross-race effect? A registered replication of Hugenberg, Miller, and Claypool (2007). *Journal of Experimental Social Psychology*, *104*, 104423.

DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00561

Desrichard, O., & Köpetz, C. (2005). A threat in the elder: The impact of task-instructions, self-efficacy and performance expectations on memory performance in the elderly. *European Journal of Social Psychology*, *35*(4), 537–552. https://doi.org/10.1002/ejsp.249

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117. https://doi.org/10.1037/0096-3445.115.2.107

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology*, *38*(6), 682-688.

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology*, *30,* 113– 125.

Dolzycka, D., Herzmann, G., Sommer, W., & Wilhelm, O. (2014). Can training enhance face cognition abilities in middle-aged adults? *PLOS ONE*, *9*(3), e90249. https://doi.org/10.1371/journal.pone.0090249

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.

Fang, F., Clapham, P. J., & Chung, K. C. (2011). A systematic review of inter-ethnic variability in facial dimensions. *Plastic and Reconstructive Surgery*, *127*(2), 874.

Felder, R. M., & Brent, R. (2009). Active learning: An introduction. *ASQ Higher Education Brief*, *2*(4), 1-5.

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, *44*(1), 155-194.

Hertzog, C., & Dixon, R.A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 227-251). The MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*(2), 215–227. https://doi.org/10.1037/0882-7974.5.2.215

Hinzman, L., Lloyd, E. P., & Maddox, K. B. (2022). The stigmatized perceiver: Exploring the implications of social stigma for cross-race face processing and memory. *Social and Personality Psychology Compass*, *16*(2), e12654. https://doi.org/10.1111/spc3.12654

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, *1*(3), 158–162. https://doi.org/10.1016/j.jarmac.2012.06.004

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

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

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, *49*(18), 2273-2284.

Innocence Project. (2020, April 15). *How eyewitness misidentification can send innocent people to prison.* https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*(1), 38-43. https://doi.org/10.1177/0956797609356508

Kouo, J. L., & Egel, A. L. (2016). The effectiveness of interventions in teaching emotion recognition to children with Autism Spectrum Disorder. *Review Journal of Autism and Developmental Disorders*, *3*(3), 254–265. https://doi.org/10.1007/s40489-016-0081-1

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77.

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.

Malpass, R.S. (1981). Training in face recognition. In G.M. Davies, H.D. Ellis, & J.W. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271-285). Academic Press.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, *14*(2), 285-292.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255-260.

McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive Neuropsychology*, *29*(1–2), 174–212. https://doi.org/10.1080/02643294.2012.660138

McKone, E., Dawel, A., Robbins, R. A., Shou, Y., Chen, N., & Crookes, K. (2023). Why the other-race effect matters: Poor recognition of other-race faces impacts everyday social interactions. *British Journal of Psychology*, *114*(S1), 230–252. https://doi.org/10.1111/bjop.12508

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, *16*(5), 778-797.

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), 3–35.

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, *33*, 783-792.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*(7), 608-615.

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

Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 509-534.

Pan, S. C. (2015). The interleaving effect: mixing it up boosts learning. *Scientific American*, *313*(2).

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams Jr, R. B., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology*, *96*(4), 795-810.

Penry, J. (1971). *Looking at faces and remembering them: A guide to facial identification.* Elek Books.

Pica, E., Warren, A. R., Ross, D. F., & Kehn, A. (2015). Choosing your words and pictures wisely: When do individuation instructions reduce the cross-race effect?. *Applied Cognitive Psychology*, *29*(3), 360-368.

R Core Team (2023). R: A language and environment for statistical computing (Version 4.3.2). R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*, 313-331.

Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology*, *100*, 717-728.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, *88*(1), 79–107. https://doi.org/10.1016/S0010-0277(03)00020-9

Simon, D., Chen, J. M., Sherman, J. W., & Calanchini, J. (2022). A recognition advantage for members of higher-status racial groups. *British Journal of Psychology, 00,* 1-24.

Singh, B., Mellinger, C., Earls, H. A., Tran, J., Bardsley, B., & Correll, J. (2022). Does cross-race contact improve cross-race face perception? A meta-analysis of the cross-race deficit and contact. *Personality and Social Psychology Bulletin*, *48*(6), 865–887. https://doi.org/10.1177/01461672211024463

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identifications.

Canadian Journal of Behavioural Science, *36*(2), 146–154. https://doi.org/10.1037/h0087225

Sporer, S. L. (2001). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy and Law*, *7*, 170-200.

Sporer, S. L. (1991). Deep—Deeper—Deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 323–333.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225-245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*, B1-B9.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 122-131.

Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M. D., & Schultz, R. T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! program. *Journal of Child Psychology and Psychiatry*, *51*(8), 944–952. https://doi.org/10.1111/j.1469-7610.2010.02258.x

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

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.

Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification: What do we do about it? *Psychology, Public Policy, and Law*, *7*(1), 230–246.

West, R. L., Bagwell, D. K., & Dark-Freudeman, A. (2007). Self-efficacy and memory aging: The impact of a memory intervention based on self-efficacy. *Aging Neuropsychology and Cognition*, *15*, 302–329. https://doi.org/10.1080/13825580701440510

Wilhelm, O., Herzmann, G., Kunina-Habenicht, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of Personality and Social Psychology*, *99*, 530–548. https://doi.org/10.1037/a0019972

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The cross-race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review*, *7*(1), 83–113. https://doi.org/10.1111/j.1751-2409.2012.01044.x

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

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Interracial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17*(3), 365-373.

Yaremych, H. E., Preacher, K. J., & Hedeker, D. (2021). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods, 28*(3), 613-630.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, *20*(2), 137–142. https://doi.org/10.1016/j.cub.2009.11.067

Appendix A

Above Chance Counts Across Experiments

	Original *N*	***N* (total *d'* > 0)**	***N* (total *d'* <= 0)**
Exp 1	320	317	3
Exp 2	196	196	0
Exp 3	255	250	5

Note. Total *d*' less than or equal to 0 indicates recognition memory at or below chance when considering all trials (i.e., across all time points and stimulus races).

Appendix B

Experiment 1 Descriptive Statistics for d', Hits, and False Alarms

	Holistic Training ($n = 150$)						Matched Control ($n = 167$)					
	White Faces			Black Faces			White Faces			Black Faces		
Time of Test	Hits	False Alarms	*d'*	Hits	False Alarms	*d'*	Hits	False Alarms	*d'*	Hits	False Alarms	*d'*
Before Manipulation	0.80^{a} (0.17)	0.15^{b} (0.13)	2.17 (0.89)	$0.83^{a,e}$ (0.15)	0.21^{b} (0.16)	2.09^{f} (0.89)	0.80 (0.19)	0.14^{g} (0.14)	2.24^{h} (0.94)	0.81^{l} (0.17)	0.22^{g} (0.18)	1.95^{h} (0.92)
After Manipulation	0.78 (0.20)	0.15^{c} (0.17)	2.16^{d} (1.11)	0.76^{e} (0.20)	0.21^{c} (0.19)	$1.85^{d,f}$ (1.06)	0.78^{i} (0.20)	0.13^{j} (0.13)	2.23^{k} (0.98)	$0.73^{i,l}$ (0.22)	0.20^{j} (0.19)	1.79^{k} (1.21)

Note. Only participants with a total *d'* above chance were included. Standard deviations within parentheses. Means with shared superscripts are reliably different from one another. Shared superscripts in the same row indicate a reliable difference at the same time point. Shared superscripts in the same column indicate a reliable difference between time points within a single condition.

Appendix C

Experiment 2 Descriptive Statistics for d', Hits, and False Alarms

	Holistic Training ($n = 94$)						Matched Control ($n = 102$)					
	White Faces			Black Faces			White Faces			Black Faces		
Time of Test	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***
Before Manipulation	$0.82^{a,p}$ (0.16)	$0.15^{b,g}$ (0.16)	2.28 (0.99)	$0.77^{a,h}$ (0.22)	0.20^{b} (0.16)	1.89 (1.05)	0.74^{p} (0.21)	0.19 (0.16)	1.87 (1.04)	0.71 (0.20)	0.22 (0.19)	1.61 (1.01)
After Manipulation	0.79^{c} (0.19)	$0.11^{d,g}$ (0.14)	2.40 (0.97)	$0.68^{c,h,i}$ (0.23)	0.17^{d} (0.17)	1.76 (1.03)	0.77^{k} (0.19)	0.15^{l} (0.18)	2.15 (1.09)	$0.70^{k,n}$ (0.23)	0.20^{l} (0.17)	1.73 (1.13)
24 Hours Later	0.75 (0.20)	$0.12^{e,q}$ (0.17)	2.24 (1.03)	$0.74^{i,j}$ (0.22)	0.19^{e} (0.20)	1.93 (1.18)	0.74 (0.20)	0.17^{q} (0.19)	1.94 (1.01)	0.76^{n} (0.20)	0.20^{o} (0.18)	1.87 (1.06)
1 Week Later	0.79 (0.20)	0.15^{f} (0.17)	2.25 (1.03)	$0.79^{j,r}$ (0.17)	0.21^{f} (0.17)	1.93 (1.03)	0.74 (0.23)	0.20^{m} (0.20)	1.89 (1.14)	0.72^{r} (0.20)	$0.24^{m,o}$ (0.20)	1.61 (1.11)

Note. Only participants with a total *d'* above chance were included. Standard deviations within parentheses. Means with shared superscripts are reliably different from one another. *d'* reliable differences are not indicated, since they are displayed in the original Experiment 2 table. Shared superscripts in the same row indicate a reliable difference at the same time point. Shared superscripts in the same column indicate a reliable difference between time points within a single condition.

Appendix D

Experiment 3 Descriptive Statistics for d', Hits, and False Alarms

	Before Manipulation						After Manipulation					
	White Faces			Black Faces			White Faces			Black Faces		
Condition	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***	**Hits**	**False Alarms**	***d'***
Holistic Training (*n* = 77)	0.81^{a} (0.15)	0.17 (0.17)	2.20 (0.91)	0.77^{b} (0.20)	0.15 (0.17)	2.13^{c} (1.06)	0.72^{a} (0.20)	0.13 (0.17)	2.03 (0.92)	$0.71^{b,v}$ (0.22)	0.16 (0.17)	1.88^{c} (0.89)
White Participants (*n* = 38)	$0.84^{d,g}$ (0.14)	0.14 (0.15)	2.42 (0.83)	$0.78^{d,h}$ (0.17)	0.14 (0.14)	2.17^{i} (0.91)	$0.76^{e,g}$ (0.16)	0.12 (0.20)	2.20^{f} (0.86)	$0.69^{e,h}$ (0.22)	0.17 (0.18)	$1.81^{f,i}$ (0.93)
Black Participants (*n* = 39)	0.79^{j} (0.16)	0.20^{k} (0.18)	1.97 (0.94)	0.77 (0.22)	0.16 (0.19)	2.09 (1.19)	$0.68^{j,x}$ (0.23)	0.14^{k} (0.14)	1.86 (0.96)	0.73 (0.21)	0.16 (0.17)	1.95 (0.84)
Individuation (*n* = 86)	0.79 (0.19)	0.17 (0.18)	2.14 (1.07)	0.77 (0.20)	0.20 (0.20)	1.95 (1.13)	0.78 (0.20)	0.17 (0.19)	2.12 (1.12)	0.79^{v} (0.19)	0.17 (0.18)	2.16 (1.14)
White Participants (*n* = 33)	0.79^{l} (0.18)	0.15 (0.14)	2.15^{m} (0.93)	0.69^{l} (0.20)	0.14 (0.12)	1.82^{m} (1.03)	0.75 (0.20)	0.15 (0.16)	2.03 (1.00)	0.78 (0.20)	0.21 (0.20)	1.92 (1.21)
Black Participants (*n* = 53)	0.80 (0.20)	0.18 (0.21)	2.14 (1.16)	0.82 (0.19)	$0.23^{n,w}$ (0.23)	2.02 (1.19)	0.80 (0.21)	0.18 (0.21)	2.18 (1.20)	0.81 (0.19)	0.15^{n} (0.16)	2.30 (1.08)
CRE Awareness (*n* = 87)	0.79 (0.19)	0.16 (0.17)	2.20^{o} (1.01)	0.76 (0.22)	0.18^{p} (0.18)	1.97^{o} (1.09)	0.77 (0.22)	0.12 (0.15)	2.27 (1.14)	0.76 (0.23)	0.14^{p} (0.17)	2.16 (1.16)
White Participants (*n* = 49)	0.82^{q} (0.17)	0.15^{t} (0.16)	2.30^{r} (0.91)	0.74^{q} (0.24)	0.20^{u} (0.19)	1.79^{r} (1.13)	0.78 (0.21)	0.11^{t} (0.14)	2.40^{s} (1.05)	0.74 (0.25)	0.14^{u} (0.18)	2.05^{s} (1.16)
Black Participants (*n* = 38)	0.76 (0.20)	0.16 (0.18)	2.08 (1.12)	0.78 (0.19)	0.15^{w} (0.16)	2.20 (1.01)	0.75^{x} (0.24)	0.15 (0.16)	2.11 (1.24)	0.79 (0.21)	0.13 (0.16)	2.31 (1.16)

Note. Only participants with a total *d'* above chance were included. Standard deviations within parentheses. Means with shared superscripts are reliably different from one another. Shared superscripts in the same row indicate a reliable difference at the same time point. Shared superscripts in the same column indicate a reliable difference between time points within a single condition.