Title
Investigating inappropriate cue utilization in the own-race bias

Permalink
https://escholarship.org/uc/item/6z96x21p

Author
Finklea, Kristin Michelle

Publication Date
2008

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Investigating Inappropriate Cue Utilization in the Own-Race Bias

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

Doctor of Philosophy

in

Psychology

by

Kristin Michelle Finklea

Committee in charge:

Professor Craig R. M. McKenzie, Chair
Professor Garrison Cottrell
Professor David Huber
Professor Tomás Jiménez
Professor Piotr Winkielman

2008
The Dissertation of Kristin Michelle Finklea is approved, and it is acceptable in quality and form for publication on microfilm:

________________________________________________________
________________________________________________________
________________________________________________________
________________________________________________________
________________________________________________________
________________________________________________________

Chair

University of California, San Diego

2008
TABLE OF CONTENTS

Signature page.................................................................................. iii
Table of contents.............................................................................. iv
List of figures.................................................................................... vi
List of tables...................................................................................... vii
List of graphs.................................................................................... viii
Acknowledgement........................................................................... ix
Vita................................................................................................... x
Abstract.......................................................................................... xii

Chapter 1: Introduction...................................................................... 1

Chapter 2: Experiment 1................................................................. 18
  Introduction.................................................................................. 18
  Method....................................................................................... 20
  Results....................................................................................... 22
  Discussion............................................................................... 26

Chapter 3: Experiment 2.................................................................. 31
  Introduction................................................................................ 31
  Experiment 2a......................................................................... 35
    Method.................................................................................. 35
    Results................................................................................ 38
    Discussion.......................................................................... 42
  Experiment 2b......................................................................... 44
    Method................................................................................ 44
    Results.............................................................................. 46
    Discussion.......................................................................... 49

Chapter 4: Experiment 3................................................................. 54
  Introduction................................................................................ 54
  Method...................................................................................... 59
  Results................................................................................... 60
  Discussion............................................................................ 68
<table>
<thead>
<tr>
<th>Chapter 5: Experiment 4</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>Method</td>
<td>79</td>
</tr>
<tr>
<td>Results</td>
<td>83</td>
</tr>
<tr>
<td>Discussion</td>
<td>88</td>
</tr>
<tr>
<td>Chapter 6: General Discussion</td>
<td>93</td>
</tr>
<tr>
<td>Appendix</td>
<td>105</td>
</tr>
<tr>
<td>References</td>
<td>123</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Experiment 1, Figure 1: Example test stimuli for Experiment 1……………… 105

Experiment 1, Figure 2: Example eyes with different iris colors. Clockwise from top left: hazel, grey, green, and blue…………………………….. 106

Experiment 1, Figure 3: The eye on the left is an example eye with a “single eyelid.” The eye on the right is the same eye, but with a “double eyelid.”………………………………………………………………………… 106

Experiment 2, Figure 1: Example test stimuli for Experiment 2. The first Asian face in the top row and the first Caucasian face in the bottom row are the original faces. The second face in each row has the “opposite eyelid” from the original face. The third face in each row has the “opposite iris color” from the original face…………………………………………………………. 107

Experiment 3, Figure 1: Example test stimuli for Experiment 3. The top row contains one Asian male and one Asian female. The bottom row contains one Caucasian male and one Caucasian female………………….. 108

Experiment 3, Figure 2: Interracial Contact Questionnaire used in Experiment 3……………………………………………………………………. 109

Experiment 3, Figure 3: Plots of the six dimensions revealed through multidimensional scaling in Experiment 3. These plots are based on Asian participants’ response data………………………………………………. 112

Experiment 3, Figure 4: Plots of the six dimensions revealed through multidimensional scaling in Experiment 3. These plots are based on Caucasian participants’ response data……………………………………… 113

Experiment 4, Figure 1: Example test stimuli for Experiment 4……………… 114
LIST OF TABLES

Experiment 1, Table 1: Dimensions mentioned in Asian and Caucasian participants’ descriptions of Asian and Caucasian test Faces........................................................................................................ 115

Experiment 3, Table 1: Correlations between dimensions from Asian and Caucasian MDS algorithms................................................................. 117

Experiment 3, Table 2: Correlations between dimensions from Asian (high and low interracial contact groups) and Caucasian MDS algorithms........................................................................................................ 118
LIST OF GRAPHS

Experiment 1, Graph 1: Caucasian participants mention iris color significantly more often than Asians when describing faces in Experiment 1. Further, all participants mention iris color significantly more when describing Caucasian faces than when describing Asian faces………………………………………………………………… …………….. 118

Experiment 1, Graph 2: Asian participants mention “single eyelids” and “double eyelids” significantly more often than Caucasians when describing faces in Experiment 1. Asians use this descriptor differentially more often for same-race Asian faces than for cross-race Caucasian faces………. 118

Experiment 2, Graph 1: Participants in Experiment 2a false alarmed significantly more on Asian faces than on Caucasian faces in the “opposite eyelid” condition, and they false alarmed significantly more on Caucasian faces than on Asian faces in the “opposite iris” condition…………….. 119

Experiment 2, Graph 2: Compared to the control condition, participants in Experiment 2a false alarmed significantly more on Asian faces in the “opposite eyelid” condition and less on Asian faces in the “opposite iris” condition. Participants also false alarmed significantly more on Caucasian faces in the “opposite iris” condition than compared to the control condition………………………………………………………………….. 119

Experiment 2, Graph 3: Caucasian participants false alarmed more on Asian faces than on Caucasian faces, and Asian participants false alarmed more on Caucasian faces than on Asian faces in Experiment 2b. All participants false alarmed equally to Asian faces with the “opposite eyelid” and to Caucasian faces with the “opposite iris” color……………………………………………………………….. 120

Experiment 4, Graph 1: Pre-training in Experiment 4, Asian participants set a higher decision criterion ($c$) for Caucasian faces than for Asian faces and Caucasians set a higher criterion for Asian faces than for Caucasian faces. This is evidence of the own-race bias……………………………….. 121

Experiment 4, Graph 2: Across conditions in Experiment 4, Asian participants set a higher decision criterion ($c$) for Caucasian faces than for Asian faces and Caucasians set a higher criterion for Asian faces than for Caucasian faces. This is evidence of the own-race bias………………………… 121
ACKNOWLEDGEMENT

I would like to acknowledge Professor Craig McKenzie and Professor David Huber for their mentorship on this dissertation. They will be co-authors on the manuscript when these experiments are submitted for publication.

I would also like to thank my family, friends, and colleagues for their advice and support.
VITA

EDUCATION

University of California, San Diego, La Jolla, California
Ph.D. in Psychology (June 2008)
M.A. in Psychology (June 2005)

University of Florida, Gainesville, Florida
B.S. in Psychology, Summa Cum Laude (December 2003)

PUBLICATIONS


TECHNICAL REPORTS


Illinois Pilot Program: Comparison of standard simultaneous lineups with sequential, blind administered lineups in three jurisdictions. Chicago Police Department, March, 2006 (with Ebbesen, E. B.)

WORK UNDER REVIEW


PAPERS IN PREPARATION


CONFERENCE PRESENTATIONS AND INVITED TALKS


“Perceptual Differences in the Own Race Bias: A Multidimensional Scaling Analysis,” UCSD All-Grad Research Symposium, February, 2008, La Jolla, CA.


“Eyewitness accuracy in the real world: DNA evidence as “ground truth” for eyewitness accuracy rates,” UCSD All-Grad Research Symposium, February, 2006, La Jolla, CA.

SERVICE

Graduate Representative, UCSD Academic Dishonesty Hearings Board (2007 – 2008)
Graduate Representative, UCSD Psychology Department Admissions Committee (2007 – 2008)
Graduate Representative, UCSD Psychology Department Colloquium Committee (2005 – 2006)
ABSTRACT OF THE DISSERTATION

Investigating Inappropriate Cue Utilization in the Own-Race Bias

by

Kristin Michelle Finklea

Doctor of Philosophy in Psychology
University of California, San Diego, 2008
Professor Craig R. M. McKenzie, Chair

The own-race (ORB) bias in face recognition is the well-known phenomenon that people are generally more accurate at recognizing and discriminating faces of their own race than any other race. The mechanisms driving the bias, however, are unclear. This series of four dissertation experiments examined the inappropriate cue utilization hypothesis in conjunction with the contact theory as the best explanation for the ORB. Experiment 1 investigated the cues that Asians and Caucasians use to describe same-race and cross-race faces. Asians mentioned the presence/absence of an eyelid crease more
did Caucasians, who more often mentioned iris color when describing faces. Further, Asians and Caucasians used these respective cues more frequently to describe same-race than cross-race faces. Experiment 2 employed two recognition memory tasks to examine whether individuals are sensitive to the natural ranges of variability for different aspects of the eye region on different races. Results imply that people are sensitive to the greater range of variability for iris color on Caucasians over Asians and to the greater range of variability for the presence/absence of eyelid creases on Asians over Caucasians. Experiment 3 used multidimensional scaling (MDS) analyses to determine the facial features that Asians and Caucasians attend to most often when judging the similarity between faces. Results indicate that Asians and Caucasians both rely most heavily on race and gender when comparing faces. Also, all participants place equal importance on iris color, while Asians place greater importance on eyelids than do Caucasians. Interracial contact analyses further show that Asians higher on interracial contact have MDS solutions that more closely resemble the Caucasian MDS solutions. Lastly, Experiment 4 trained participants to attend to features appropriate for cross-race face discrimination in attempt to mitigate the ORB. Although the training did not affect participants’ ORB, attending participants to the race of faces did increase the discriminability of all faces. These four experiments imply that the ORB may be driven by individuals lacking the experience with attending to those cues most appropriate for cross-race discrimination.
CHAPTER 1

Introduction

In both academic and applied domains, attention has recently turned to eyewitness identifications, particularly the potential role they play in wrongful convictions. According to the Innocence Project (2001), 75% of exoneration cases include some form of mistaken eyewitness identification of the defendant. Although it is not known whether this mistaken ID was the primary cause of the conviction, it is nonetheless alarming that there exists such a high rate of mistaken identification involved in the conviction of innocent suspects. The spotlight then shifts to the factors affecting an eyewitness’s memory of a culprit. One such factor concerns whether a witness is attempting to identify a suspect of the same or different race. It is a robust finding that people are often more accurate in recognizing and discriminating between members of their own race than recognizing individuals of a different race (e.g. Feingold, 1914; Brigham & Malpass, 1985; Anthony, Copper, & Mullen, 1992; Meissner & Brigham, 2001; Sporer, 2001). This phenomenon is known as the own-race bias (ORB) or cross-race effect, and is considered generally reliable across races (Kassin, Ellsworth, & Smith, 1989).

In this paper, recognition accuracy is defined in terms of hits (correct identification of a target or acknowledgment that a target has been previously viewed) and false alarms (incorrect identification of a non-target or falsely indicating that a target has been previously viewed). Many researchers use discrimination accuracy ($d'$), which is the difference between the z-score of hits and the z-score of false alarms, to evaluate participants recognition accuracy for same- versus cross-race faces (Meissner &
Brigham, 2001). Also, most research examining the ORB employs face recognition studies rather than event memory paradigms. In a typical face recognition task, a subject is shown a number of faces (e.g. 40) and tested on a larger set of faces (e.g. 80) consisting of a combination of original faces and novel stimuli; participants then indicate whether a given test face is “old” or “new.” For the present discussion, a racial category is defined by the referent’s expressed phenotype rather than genotype.

Although the ORB is well-established, the field lacks a unified theory for the psychological mechanisms underlying it (e.g. Meissner & Brigham, 2001; Sporer, 2001; Ayuk, 1990). Consequently, the goals of the present paper are two-fold: 1) examine the present theories in the face recognition and eyewitness identification literature surrounding the ORB and discuss the elements needed for a complete theoretical understanding of the effect and 2) report experiments that support a combination of theories as an explanation for the own-race bias.

**Review of Current Theories**

Multiple theories have been proposed to explain the ORB (Ayuk, 1990; Brigham & Malpass, 1985; Shepherd, 1981), but they are not mutually exclusive.

**A) Inherent difficulty hypothesis**

One of the first theories behind the ORB suggested that there is an intrinsic obstacle involved in discriminating between members of some racial groups more so than others, independent of witness race (Brigham & Malpass, 1985). However, this hypothesis is counter to the ORB in that an inherent difficulty would make faces of a particular race more difficult to differentiate not only for outgroup members, but for ingroup members as well. The theory suggests, for example, that a given race X may be
more physically homogenous than a given race Y. The prediction would then follow that both members of race X as well as members of race Y should have more difficulty discriminating between members of race X than race Y. Without analyzing any evidence, this theory can only provide uni-directional support for any study examining the cross-race effect. Any theory that is to provide the desired explanation for the ORB must receive bi-directional support such that race X is more accurate in discriminating between other members of race X than race Y while a person of race Y is better at discriminating between other persons of race Y than race X.

This inherent difficulty theory has received inconsistent support, with some researchers finding white faces more easily recognizable than black faces (Malpass & Kravitz, 1969; Cross, Cross, & Daly, 1971), some supporting the ease of recognition for black over white faces (Brigham & Barkowitz, 1978; Brigham & Williamson, 1979; Ellis & Deregowski, 1981), and some finding no disparity in ease of recognition for any one racial group over another (Shepherd, Deregowski, & Ellis, 1974; Chance et al., 1975; Feinman & Entwistle, 1976). Without testing for homogeneity of black and white stimulus faces used in their experiment, Malpass and Kravitz (1969) found that white stimuli were recognized more often than black stimuli by both white and black participants. Specifically, their results only demonstrated a consistent ORB for white participants, but not for black participants. The results may have been moderated by amount of interracial contact such that blacks with greater experience with white faces demonstrated better discrimination of white faces; perhaps the whites assessed as having increased interracial contact still had significantly less interracial contact than the blacks assessed as being high-contact and thus did not exhibit better discrimination of black
faces. Divided results were also found by Cross et al. (1971) wherein whites recognized white faces more often than black faces, but blacks showed no difference in overall recognition between white and black faces (both blacks and whites had more false alarms on black faces than on white faces).

Brigham and Barkowitz (1978), like others, did not analyze the homogeneity of photos used in their research, but did find a main effect of target race in that blacks were recognized with greater ease than whites, opposite of the findings of Malpass and Kravitz (1969). Contrary to the basic premise of the inherent difficulty hypothesis, however, a complete ORB was observed for both black and white subjects in the study. Similar results were obtained by Chance et al. (1975), finding a full ORB for black and white participants tested on black, white, and Japanese faces. No main effect of race was reported, further evidence against the notion that greater homogeneity of one race over another may result in the ORB. Without analyzing the homogeneity of the test faces used in the referenced studies, no definitive conclusions can be drawn (Shepherd, 1981).

Studies producing mixed results on the recognition ability of same-race and cross-race faces (some showing main effects of race supporting greater discrimination of blacks, main effects supporting greater discrimination of whites, and some lacking any main effect of race) suggest that the ORB is not explained by the inherent difficulty of identifying a particular race.

B) Prejudicial attitudes hypothesis

Another preliminary theory behind the own-race bias contends that prejudicial attitudes toward outgroup members are correlated with poor recognition/identification of those outgroup members (Seeleman, 1940; Brigham & Malpass, 1985). Seeleman (1940)
found that white participants holding positive attitudes toward blacks correctly recognized black faces more so than did those white participants who held strong negative attitudes toward blacks. If racial attitude alone is the driving force behind the ORB, then only those members of race X who have prejudicial attitudes toward race Y will demonstrate the ORB; likewise, researchers should fail to find this cross-race recognition deficit among those members of race X who do not hold prejudicial attitudes toward race Y.

Brigham and Barkowitz (1978) examined prejudicial attitudes and, while participants significantly demonstrated the ORB, they found no effect of attitude on face memory for cross-race faces. In one of the few event-memory studies examining the ORB, Aronstam and Tyson (1980) simulated assaults (both whites assaulting blacks as well as blacks assaulting whites) in South Africa, as this location was deemed to be a particularly race-conscious society. It was hypothesized that racial biases may negatively influence subjects’ resulting cross-race descriptions of the events and perpetrators. Attitudes, however, did not significantly impinge upon the reliability of witness descriptions of either the event or the perpetrator. Lavrakas, Buri, and Mayzner (1976) tested white subjects’ face memory for black faces and found no correlation between reported racial attitude and discrimination ability; instead, a small association was found between subjects’ prior experience with blacks and their recognition accuracy for other-race faces. Similarly, Platz and Hosch (1988) found a small correlation between recognition accuracy and attitudes towards Mexican-Americans for the black participants in their study, but this correlation disappeared when the level of interracial contact was controlled for.
In short, studies have consistently failed to establish much support for a direct link between racial attitudes and discrimination ability for other-race faces. The missing element correlating with both racial attitude and memory for other-race faces seems to be the extent to which an individual has had contact with members of other racial categories (Lavrakas et al., 1976; Brigham & Barkowitz, 1978; Brigham & Ready, 1985; Meissner & Brigham, 2001). Brigham and Barkowitz (1978) noted this correlation only existed for white subjects in their sample, but not for blacks. One example of this link between attitude and interracial contact was presented by Chance and Goldstein (1996) such that in examining post-desegregation and post-civil rights eras we may see an increase in the amount of interracial contact and perhaps a subsequent decrease in cross-race prejudicial attitudes. If attitude alone were affecting the ORB, researchers would see a decline in the magnitude of the effect since the 1960s. Instead, the magnitude of the ORB has remained with a decreased influence of attitude on cross-race recognition and a mirrored increased influence of interracial contact on recognition accuracy (Meissner & Brigham, 2001).

C) Inappropriate cue utilization hypothesis

The inappropriate cue utilization hypothesis suggests that the divergence in same-race and cross-race identification accuracy may arise because people use certain cues that are effective in distinguishing between members of one’s own racial category; these cues may not be appropriate for distinguishing between members of a different racial category but are nonetheless employed (Shepherd, 1981). Such a theory would suggest that we do not attend to (and possibly do not encode) those features which distinguish one race from another, but rather we attend to features which help us make more individuating distinctions among other members of our own race.
The theory would hypothesize that someone from racial category X may use characteristics (a), (b), and (c) to discriminate between other members of race X and would likewise use these same three characteristics to discriminate between members of racial category Y. The suggested problem underlying the own-race bias is that perhaps characteristics (c), (d), and (e) should actually be engaged to best discriminate between members of race Y demonstrated by the fact that members of race Y use cues (c), (d), and (e) to discriminate own-race faces.

Evidence supporting elements of the inappropriate cue utilization hypothesis comes primarily from descriptions given by Black Africans and White British subjects to pictures of same-race (though not cross-race as well) faces (Ellis, Deregowski, & Shepherd, 1975). White participants more often used descriptors for hair texture, hair color and eye/iris color, while the Black participants more often mentioned hair style, face shape, chin, ears, eyebrows, and whites of eye. Similar results were noted by Doty (1998) across races, with Caucasian subjects focusing on hair and non-Caucasians focusing on face shape, eyes, lips, and nose. Sporer (2001) suggests that given a facial description (either cross- or same-race), if the frequency with which each feature is mentioned is equated with the amount of attention given to the particular feature, we may be able to explain the superiority of own-race recognition. Unfortunately, researchers have yet to determine that the features attended to most often by members of a particular race are actually those features which are most useful in discriminating between members of that race.

Additionally, it is possible that some features should be weighted more heavily than others. This question has not been addressed by the inappropriate cue utilization
theory itself or by researchers evaluating the theory. Following the notion of feature utility in terms of discrimination accuracy, Clifford and Bull (1978) as well as Sporer (2001) indicated that people may infer less memorable characteristics of an individual on the basis of other well-remembered characteristics (e.g., inferring blonde hair, given a memory of blue eyes). A witness will be accurate so long as those paired characteristics actually correlate in the given instance. Of critical importance would be determining if members of one race understand the extent to which common features correlate in other-race categories or if this assumption is based on the expected rate of correlation within their own race.

Essentially, the inappropriate cue utilization implies that there may be two distinct issues at hand: 1) people may inappropriate apply cues used to discriminate own-race faces to the discrimination of other-race faces and/or 2) people may fail to use those cues appropriate for other-race face discrimination. It may be that people use inappropriate cues at the expense of appropriate cues for other-race face recognition or it may be that people simply are not using the most appropriate cues. This is another distinction that the inappropriate cue utilization hypothesis has not clarified.

D) Differential processing hypothesis

In the same perceptual learning domain as the inappropriate cue utilization hypothesis lies the theory of differential processing. The notion of differential processing has at its foundation the idea that there is a hierarchy of cognitive levels in memory where information is processed (Craik & Lockhart, 1972). Craik and Lockhart suggest that familiar stimuli “will be processed to a deep level more rapidly than less meaningful stimuli and will be well-retained” (p. 676). Presumably, faces from one’s own race are
more familiar than faces from a different race. The depth of processing theory would hypothesize that given a person of race X, he will process other faces of race X at a deeper level during the encoding phase of memory than will he process faces of race Y.

With respect to the ORB, eyewitness researchers have not defined what is meant by a “deeper level” of memory encoding. Does this mean more information or a different type of information? For one, Levin (1996, 2000) proposed that cross-race faces are encoded with more race-specifying information and less individuating information while same-race faces are encoded with more individuating information. Specifically, he suggests that the ORB is influenced by people perceptually encoding race as a visual feature for cross-race faces (though not for same-race faces) and thus reducing the amount of individuating information available to encode.

Several researchers have investigated the idea that people use different orienting strategies for same- and other-race faces; moreover, people may process same-race faces in a more subjective, inferential manner (focusing on internal (e.g. friendly, honest) characteristics and qualities) while processing cross-race faces in a more physical, superficial manner (focusing on actual external facial attributes such as eye color or nose size) (Chance & Goldstein, 1981; Devine & Malpass, 1985). Chance and Goldstein (1981) found white subjects making inferential, thus deeper, more subjective descriptions of same-race faces while making physical descriptions of other-race faces. If any differences in same- and cross-race recognition ability stemmed from differences in depth of processing (defined by depth/quantity of face description), equating these descriptions should eliminate any observed cross-race effects. For example, Devine and Malpass (1985) focused half their subjects on superficial aspects of a face (Black versus White).
and half their subjects on inferential qualities of faces (friendly versus unfriendly). Irrespective of condition, an ORB was observed among both White and Black participants, though both same- and cross-race recognition decreased in accuracy in the condition wherein subjects were asked to focus on inferential qualities of faces. Perhaps encoding a face inferentially rather than superficially simply detracts from encoding actual physical features, thus making subsequent recognition inherently more difficult. Devine and Malpass (1985) concluded that the differential processing theory could not account for the ORB. If “depth of processing” were examined in terms of more specific information or a greater amount of information rather than different information, researchers may garner different results which could provide a more thorough understanding of the mechanisms underlying the ORB.

E) Contact hypothesis

The most widely supported theory behind the own-race bias is the contact hypothesis, which contends that the greater amount of interracial contact one has, the more accurate he will be at recognizing and discriminating among faces from a different racial category (Brigham & Malpass, 1985). In this basic form, the theory would predict a negative correlation between amount of interracial contact and the ORB. In this form, the theory does not discuss the amount or quality of contact needed to mitigate the ORB effects, nor does it discuss how this amount and/or quality of contact would be measured. Ng and Lindsay (1994) proposed that shallow contact may not dissipate the ORB as much as would closer, more personal contact, an idea based on Craik & Lockhart’s (1972) notion concerning depth of processing.
The evidence for the contact theory has been mixed, with some researchers finding a negative correlation between quantity of interracial contact and subsequent recognition/discrimination accuracy (Chance et al., 1975; Brigham, Maass, Snyder, & Spaulding, 1982), and others finding no correlation (Malpass & Kravitz, 1969; Luce, 1974; Brigham & Barkowitz, 1978). This suggests that the contact theory, in this general form, cannot explain the ORB. Furthermore, the lack of support for the theory cannot result directly from a flaw in measurement techniques, as conflicting findings result irrespective of the means by which level of contact is determined. For example, Chance et al. (1975) claimed that Black and White participants in their subject pool would have greater contact with one another than either race would have with Japanese faces. In line with this hypothesis, although both Blacks and Whites demonstrated the ORB (with greater accuracy on same-race faces than on either of the two other-race face groups), both also displayed superior recognition for both Blacks and Whites than for Japanese faces. Ng and Lindsay (1994) measured contact similarly, primarily by the cross-cultural nature of their sample, but found contradictory results: the level of interracial contact did not significantly affect recognition performance for either the Canadian or Singaporean subjects.

Inconsistent support for the contact hypothesis has also been garnered using self-report techniques to assess the level of contact subjects have with other racial groups. In a typical face recognition task, Slone, Brigham, and Meissner (2000) obtained a significant negative correlation between amount of self-reported recent interracial contact and the white participants’ tendencies to exhibit the ORB. On the other hand, several other research teams have not found such a relationship using a self-report measure of
interracial contact (e.g. Malpass & Kravitz, 1969; Cross et al., 1971; Brigham & Barkowitz, 1978). Brigham and Barkowitz (1978) measured interracial contact by 11 individual items, and each had a near-zero correlation with cross-race recognition accuracy.

Field studies have also examined the correlation between interracial contact and recognition accuracy and have yielded inconsistent results as well. Brigham et al. (1982) found that white clerks’ recognition ability for black faces varied in accordance with the amount of self-reported cross-race experience. However, experience and recognition ability did not correlate among the Black store clerks. Platz and Hosch (1988) used a methodology closely paralleling Brigham et al. (1982), but found results which contradicted their predecessors. Anglo, Black, and Mexican-American clerks all demonstrated an own race bias. Although white clerks had more experience with Mexicans than with Blacks (and blacks had more experience with Mexicans than with whites) they identified Mexican and Black customers equally well (and Black clerks demonstrated superior identification of White than Mexican customers). Only the correlation between Black clerks’ experience levels with Mexicans and their subsequent recognition of Mexican customers was clearly in support of the contact hypothesis. To wholly explain the ORB, the contact theory would hold across all races. However, the theory has received most support under conditions having White witnesses and Black targets, thus preventing full generalization of the theory (Brigham et al., 1982; Kleider & Goldinger, 2001; Wells & Olson, 2001).

How exactly does interracial contact affect one’s memory of individuals from a different racial category? Quality of contact has been discussed as one potential
moderator for the interracial contact hypothesis (Brigham & Malpass, 1985).

Investigators have gathered information on the level of interracial contact based on population statistics (e.g. Aronstam & Tyson, 1980; Platz & Hosch, 1988) as well as self-reports on the number of cross-race associations (e.g. Malpass & Kravitz, 1969) or frequency of interracial contact (e.g. Brigham & Barkowitz, 1978). Without further specificity within the contact theory, researchers cannot determine what, if any, aspect of interracial contact may breed more accurate recognition of other-race persons. Malpass and Kravitz (1969) long ago suggested that “experience occurring when physiognomic discriminations are of instrumental value are more important than other forms, such as ‘mere’ exposure (p.334).” What they do not suggest, however, are the specifics surrounding the physiognomic discriminations between other-race persons. For instance, how are these discriminations between other-race persons made? Are these discriminations made along the same dimensions as those made between individuals of one’s own race? How does one learn the appropriate cues to make these other-race physiognomic discriminations? These questions may be answered by assimilating the contact theory with other, already-existing theories of the ORB (i.e. inappropriate cue utilization).

Understanding the ORB

From reviewing the current theories behind the ORB, it is evident that the basic concept of the contact theory must be true (e.g. Levine, 2000). For this theory to be false, quantity of interracial contact would never have an effect on recognizing cross-race faces; some races would necessarily be more physiologically homogenous than others, and researchers have found no support for this (e.g. Shepherd, Deregowski, & Ellis, 1974;
Chance et al., 1975; Feinman & Eintwistle, 1976). Further, support is only found for the prejudicial attitudes hypothesis when it is framed in terms of interracial contact; people who hold prejudicial attitudes have less interracial contact, and when contact is taken into account, prejudicial attitudes alone do not influence the ORB. Research indicates that the amount of (or lack of) contact alone, however, is not driving the ORB (e.g. Levine, 2000; Malpass & Kravitz, 1969), and that the contact theory may be best understood in combination with the inappropriate cue utilization hypothesis.

It is proposed that the ORB may be conceptualized in the following manner:
Through contact, individuals learn to recognize and discriminate between faces. Individuals develop expertise with exposure, and thus have more expertise with faces of their own race than of a different race. Expertise entails learning the appropriate cues necessary for discrimination, opening the doors to differentiations between members of one’s own race. Unless this same expertise is developed with faces of other races, those cross-race faces will not be distinguished from one another using the appropriate cues, thus preventing recognition between members of another race. None of these concepts are alone novel; rather, they simply have yet to be assimilated within the literature and developed into a complete story to explain the ORB.

Overview of Experiments

In this paper, we discuss four experiments that advance the theoretical framework surrounding the ORB as well as introduce novel methodologies used to examine the bias. These studies examine Asians and Caucasians across different perceptual and memorial tasks in order to test the inappropriate cue utilization theory and the contact hypothesis of the ORB. Specifically, the inappropriate cue utilization theory currently does not specify:
1) whether participants are aware of the different cues they use to distinguish own- and other-race faces, 2) whether the same cues are used in perception and in memory, and 3) how readily the appropriate cues for discriminating a particular race can be used (i.e. how do learning and the contact theory specifically relate to the inappropriate cue utilization theory). We address these questions through the four experiments presented.

In Experiment 1, we asked participants to describe faces in order to elucidate those cues that people spontaneously mention when looking at same-race and cross-race faces. We used Asian and Caucasian participants in all our experiments primarily because those are the two most prominent races at the University of California, San Diego. Further, research to date has not examined the difference in face descriptions between Asian and Caucasian viewers to understand the importance of different facial cues to these two races. In particular, we found that Asians describe the presence/absence of an eyelid crease in the upper eyelid more than do Caucasians (and disproportionately more for same-race faces) and Caucasians describe the color of the eyes more than do Asians (also disproportionately more for same-race faces). Identifying cues that are mentioned more by a given race provides evidence for the inappropriate cue utilization hypothesis behind the ORB, and it suggests that those cues may vary significantly within this race and thus may be most diagnostic for discrimination.

In Experiment 2, we used a standard face recognition paradigm to measure participants’ sensitivity to changes in facial features revealed by Experiment 1 that have either a naturally large or small range of variability on either Asian or Caucasian faces. Features with a high range of variability for one race but a low range of variability for another race (e.g. eyelid crease varies more among Asians than Caucasians and iris color
varies more among Caucasians than Asians) should be differentially more diagnostic for the race with greater variability in this cue.

In Experiment 3, we used multidimensional scaling (MDS) techniques to address whether people perceptually attend to those features that appear to be most diagnostic for same-race face discrimination based on the cues that participants described in Experiment 1. Researchers have not examined via MDS the relative importance of various features to two different races of viewers. We found that although Asians and Caucasians use some similar dimensions when comparing faces, they place differential importance on these cues. We add to the current understanding of the inappropriate cue utilization hypothesis by showing that cues should not be thought of as strictly appropriate or inappropriate, but rather on a continuum of appropriateness for a given race. Furthermore, we demonstrate the importance of interracial contact in learning the appropriate cues for discrimination; our results officially assimilate two theories (contact and inappropriate cue utilization) that have only been assimilated in memory tasks, but not yet assimilated in a perceptual task.

In Experiment 4, we implemented a training task to determine whether short-term training on cues appropriate for cross-race face discrimination (but not relevant to same-race discrimination) can mitigate peoples’ demonstration of the ORB. Specifically, we trained participants on presence/absence of eyelid crease, a cue more diagnostic for discriminating Asian than Caucasian faces, to see whether this training would improve Caucasians’ recognition accuracy for Asian faces. Results imply that although people may be easily attuned to the differences between races, short-term cue training is insufficient to moderate the ORB.
These four experiments contribute to the current literature in two important ways. Firstly, we clarify the inappropriate cue utilization hypothesis by discussing the theory not only in terms of inappropriately using the wrong cues for cross-face discrimination but failing to learn (through contact) and use the appropriate cues. We further refine the theory by specifying the degree to which people apply same-race face cues to the discrimination of cross-race faces. Secondly, we employ innovative methodologies for analyzing both the perceptual and memory basis of the own-race bias.
CHAPTER 2

Experiment 1

Introduction

Because we are interested in learning the appropriate cues to use when discriminating between faces of a given race, we sought to learn which cues “experts” use. In other words, which cues do Asians use when recognizing Asian faces, which cues do Caucasians use when recognizing Caucasian faces, etc. The first step in identifying the appropriate cues for a given race is to identify those cues with a wide range of variability for one race, but not for another race. Secondly, it is important to determine if indeed face experts (but not face novices) of a given race use these highly diagnostic cues when discriminating and recognizing faces of that race.

This notion is very similar to Levin’s (1996, 2000) feature selection model which supports the idea that face processing involves selecting and processing the most diagnostic visual features. The feature selection model contends that diagnostic features for same-race faces contain individuating information whereas the diagnostic information for cross-race faces is the “race feature.” Specifically, the race feature is a feature coded by cross-race observers, but not by same-race observers, and it contains race-specifying information. Levin (1996) tested this theory using a visual search task and found that White participants searched more quickly for a cross-race Black face within a group of White faces than they searched for a same-race White face within a group of Black faces. In essence, Levin contends that processing the race feature prevents other individuating information from being encoded and processed. The feature selection model explains the
own-race bias in terms of individuation such that if individuals were able to process
diagnostic features pertinent to individuation in cross-race faces, these individuals would
not exhibit the ORB. What this model does not specify, however, are those features that
contain the individuating information in same-race faces. We are left wondering what
features Caucasians use to individuate other Caucasians, what features Asians use to
individuate other Asians, etc.

Researchers have attempted to gather this information with several different
techniques, and one of the most popular methods is a face description task. For example,
Ellis, Deregowski, and Shepherd (1975) asked participants to describe a face in detail so
that a friend could pick this face out of a crowd at a nearby train station. They examined
the frequencies with which both Whites and Blacks described different facial features.
Whites referred to iris color, hair color, and hair texture more often than did Black
participants, and Blacks referred to face outline, hair position, ears, chin, eye size, and
eyebrows more often than did White participants. Doty’s (1998) recognition experiment
also asked participants to identify those facial characteristics that are most useful for face
recognition. Caucasians reported that hair, face shape, and skin tone were the most useful
characteristics while non-Caucasian subjects (race unreported) mentioned face shape,
eyes, lips, and nose as useful facial features.

Face description tasks allow researchers to measure those features to which
participants spontaneously attend. This method investigates the cognitive strategies that
individuals may use when evaluating faces and extracting the most important
information. To date, researchers have not yet used this explicit measure of face
descriptions either: 1) to gather information on Asian faces or 2) to compare descriptions
of Caucasian faces to those of Asian faces. The present study collects explicit face
description information from both Asian and Caucasian participants in response to
viewing Asian and Caucasian faces. This is the necessary first step before later
investigating how these self-reported cues relate to memory abilities and perceptual
similarity judgments to faces of the studied races.

In this first experiment we explore what descriptors Asian and Caucasian
participants (chosen based on demographic availability) use to describe faces and
whether participants describe faces with cues that are more variable for faces of their own
race (and thus presumed to be more diagnostic) more often than they point out less
variable cues. Levin’s (2000) feature selection model would predict that both Asians and
Caucasians will mention the race of the test face for cross-race faces more often than for
same-race faces. Additionally, the inappropriate cue utilization hypothesis would predict
more specifically that Caucasians will mention those cues diagnostic for discriminating
Caucasian faces when describing both Caucasians and Asians and that Asians will refer
to those cues diagnostic for discriminating other Asians when describing both Asians and
Caucasians. This theory does not currently make any predictions about whether
participants will apply race-appropriate descriptors disproportionately more to same-race
faces than cross-race faces or whether participants will apply these descriptors with equal
frequency across all races. The most basic form of the theory implies that people will
mention race-appropriate cues with equal frequency to all faces.

Method

Participants
All participants (N = 134) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 21% male, 42% Caucasian, and 58% Asian.

Stimuli

The stimuli consisted of 40 pictures (300 x 400 pixels, gray scale): 10 Caucasian men, 10 Caucasian women, 10 Chinese men, and 10 Chinese women. The original photographs were provided by Dr. Kang Lee1. These pictures were taken under similar lighting and background. All faces had neutral expressions and contained no facial hair or jewelry. Each face was placed inside a neutral oval so that only the face and hair directly around the face were visible to the participants (see Experiment 1, Figure 1 for examples of the test stimuli). Additionally, participant responses were collected on the keyboard using the entire set of keys. Stimuli presentation as well as participant response data were controlled and collected by an E-Prime script.

Procedure

Participants viewed eight of the 40 faces. The eight faces were randomly selected (with the following constraints: two Asian men, two Asian women, two Caucasian men, and two Caucasian women) from the set of 40 for each participant and were presented in random order. Because the faces were randomly selected and presented to participants, participant responses were not constrained to a set of faces with pre-determined features; rather, responses were made to faces representative of the population. Further, across subjects, the experiment obtained ratings for all 40 faces. Participants were asked to

1 Dr. Kang Lee, Professor and Director, Institute of Child Study, University of Toronto, 45 Walmer Road, Toronto, Ontario, Canada M54 2X2, kang.lee@utoronto.ca
provide a written description for each of the eight faces based on the following instructions: “Pretend that you are describing this face to a friend. Your friend should be able to recognize this face in a crowd on the basis of your physical description.” The experiment was self-paced, and participants were allowed as much time and response space as needed to provide their descriptions.

After completing this computer-based task, participants completed a brief survey of personal demographic information including zip codes of where participants were born and of where they have lived the longest in life. Results from an extensive interracial contact questionnaire used in a pilot study as well as in Experiment 3 (to be reported later) revealed that the one item predictive of interracial contact (for individuals in our sample) was the zip code of the area where an individual had lived the longest in his or her life. Therefore, this was the only section from the questionnaire that was given to participants in this experiment.

Results

For each of the eight faces that a participant viewed, the description was coded into 35 potential face dimensions. See Experiment 1, Table 1 for a list of all 35 dimensions, along with the frequencies by which Asians and Caucasians mentioned each dimension when describing both Asian and Caucasian test faces. We were then able to analyze the differences between the amount and type of descriptors that both Asian and Caucasian participants provided for faces of their own race and for faces of the other race.

Although the range of descriptors that participants used varied across 35 dimensions, each participant used far less than this total amount (on average) when
describing a face. Collapsing across test face race, Caucasian participants ($M = 7.56$) used significantly more unique descriptors than did Asian participants ($M = 7.05$), $F(1,132) = 14.24, p < .001$. There was no interaction between participant race and the race of the test face on the number of descriptors provided. When describing a face, 80% of Asians mentioned the race of a face, while significantly fewer Caucasians (74%) mentioned race, $F(1,132) = 6.27, p = .01$. Also within the domain of race descriptors, 86% of all participants referenced the race of Asian faces, which is significantly more than the 69% of participants who mentioned the race of Caucasian faces, $F(1,132) = 43.32, p < .001$. Also collapsing across test face race, Asians mentioned nose shape and eyelids significantly more often than Caucasians, while Caucasians mentioned the nose width, nose length, eyebrow color, iris color, skin color, lip size, and ear size more than Asians. This may be considered evidence of inappropriate cue utilization if the features that a given race describes are more useful and appropriate for discriminating same-race faces. Based on these results as a function of participant race while collapsing across test face race, we performed additional analyses to examine how these differences in participant race related to test face race. This revealed for both iris color and the presence/absence of eyelid crease, a significant interaction between participant race and test face race on the frequency with which the descriptor was mentioned. These were the only two dimensions for which there was a significant main effect of participant race and a significant interaction between participant race and test face race.

With respect to participants’ descriptions of iris color (see Experiment 1, Figure 2 for example faces with different iris colors), we found significant main effects of participant race as well as test face race on the frequency with which this descriptor was
mentioned. Caucasian participants (M = .43) mentioned iris color more frequently than did Asian participants (M = .18), F(1,132) = 84.89, p < .001, and all participants mentioned iris color with respect to Caucasian test faces (M = .39) more than to Asian test faces (M = .23), F(1,132) = 34.05, p < .001. In addition to the main effects, there was a marginally significant interaction between participant race and test face race with regard to the number of iris color descriptors, F(1) = 3.62, p = .057. Caucasians mentioned iris color far more often for Caucasian faces than for Asian faces, and although Asians did mention iris color more for Caucasians than Asians as well, this disparity was not nearly as great as the difference in Caucasian participants’ descriptions (see Experiment 1, Graph 1). This is evidence of inappropriate cue utilization in that the prevalence of iris color in Caucasian descriptions indicates this is an appropriate cue to use for Caucasian discrimination, but Asians fail to use iris color (to the same extent as Caucasians) in their description of Caucasian faces. Further, Caucasians inappropriately apply iris color descriptors to Asians faces, when this may not be an appropriate cue to use for Asian face discrimination, indicated by the shortage of Asians mentioning iris color in their descriptions.

Of other interest was the reference by Asian participants to a face having either “single eyelids” or “double eyelids.” This language refers to the presence or absence of a crease in the upper eyelid which visibly divides the eyelid into two distinct sections. An eyelid without a visible crease in the upper lid is considered a “single eyelid,” while an eyelid with a visible crease is deemed a “double eyelid” (see Experiment 1, Figure 3 for example faces with single and double eyelids). Remarkably, not a single Caucasian participant made reference to this eyelid crease, but there were 53 cases of Asian
participants mentioning this descriptor. There were significant main effects of participant race, $F(1,132) = 44.38, p < .001$, as well as study face, $F(1,132) = 37.88, p < .001$, on the frequency with which eyelid crease descriptors were mentioned. Asian participants ($M = .09$) referenced eyelids more often than did Caucasian participants ($M = .00$), and participants (only Asians in this case) mentioned eyelids on Asian test faces ($M = .08$) more often than on Caucasian test faces ($M = .003$). We also found a significant interaction between participant race and study face race on the number of eyelid crease descriptors, $F(1,132) = 37.88, p < .001$ (see Experiment 1, Graph 2). Of the 53 descriptions by Asian participants that indicated the presence or absence of an eyelid crease, only two of those descriptions were of Caucasian study faces. This is also indicative of inappropriate cue utilization (or more accurately, lack of appropriate cue utilization) in that the prevalence of eyelid descriptors in Asian descriptions implies this is an appropriate cue to use for discriminating Asian faces, but Caucasians fail to use this cue at all when describing Asian faces.

**Interracial Contact**

We investigated the effect of interracial contact on the descriptions that participants gave of the Asian and Caucasian test faces. If an Asian participant lived the longest in an area where the proportion of Caucasians outweighed the proportion of Asians, this participant was coded as a high-contact participant ($N = 43$). Otherwise, the participant was considered low-contact ($N = 35$). Because only two Caucasians in our sample had lived the longest in areas where the proportion of Asians outweighed the proportion of Caucasians, we focus our interracial contact analyses solely on the Asian participants.
Of interest, we found no significant differences between the high-contact Asians and the low-contact Asians. Specifically, the high-contact Asians ($M = 7.10$) did not differ from the low-contact Asians ($M = 6.98$) on the average total number of descriptors used to describe a face, $t(76) = .67, p > .05$. Recall that we found a significant interaction between participant race and test face race on the frequency of mentioning iris color and presence/absence of an eyelid crease. Investigating whether interracial contact influenced these two descriptors, the high-contact and low-contact Asians did not differ on the frequency with which they mentioned the presence or absence of an eyelid crease ($t(76) = .33, p > .05$) or iris color ($t(76) = 1.50, p > .05$) in their descriptions. High-contact Asians ($M = .08$) mentioned eyelid creases equally frequently as low-contact Asians ($M = .09$), and high-contact Asians ($M = .20$) mentioned iris color equally frequently as low-contact Asians ($M = .16$). Interestingly, the trend was such that high-contact Asians did mention iris color more than low-contact Asians suggesting that Asians with greater contact with Caucasians may learn the importance of iris color as a discriminating cue for faces. Furthermore, there was no effect of interracial contact on the frequency of mentioning any of the other cues for which we found an overall main effect of participant race (test face race, nose shape, nose width, nose length, eyebrow color, skin color, lip size, and ear size).

Discussion

We were interested in determining those cues that different races are aware of measured by their self-reported descriptions of Asian and Caucasian faces. Specifically, participants were asked to describe Asian and Caucasian faces in order to elucidate those features that people spontaneously attend to in same- and cross-race faces.
Inappropriate Cue Utilization Hypothesis

We found significant differences in the cues that Asians and Caucasians generate when describing faces. Asians mentioned nose shape and eyelids more often than Caucasians, who mentioned the nose width, nose length, eyebrow color, iris color, skin color, lip size, and ear size more often. The inappropriate cue utilization hypothesis implies that the cues that people use to discriminate faces are those cues most important for same-race face discrimination. It follows that these cues may be differentially important for discriminating Asian and Caucasian faces. For one feature mentioned most by each race of participant, there was a significant interaction between participant race and test face race on the frequency with which the feature was mentioned. For Caucasians, this feature was iris color, and for Asians, this feature was presence/absence of an eyelid crease. These interactions are important to note because of what they potentially tell us about cue utilization. A main effect simply indicates that one race uses the cue more than another race. An interaction, on the other hand, implies that participants using these features for same-race discrimination may have some understanding that these cues are differentially more important for same-race faces, even if they apply these cues to other-race faces. We will discuss the implications for the inappropriate cue utilization hypothesis with specific reference to iris color and presence/absence of an eyelid crease.

Recall that the inappropriate cue utilization hypothesis predicted that all participants would reference features most diagnostic for same-race face discrimination. In line with this prediction, the ideal observer model (e.g. Geisler, 1989) suggests that people use those cues with the most variability (and thus increased diagnosticity) to
discriminate between faces. It follows that iris color may be more diagnostic for recognizing and discriminating Caucasian faces while the presence/absence of an eyelid crease may be most diagnostic for Asians. Although Caucasians mentioned iris color more often than Asians, and they mentioned it with reference to both same-race and other-race faces, they pointed out this feature significantly more often when describing Caucasian than Asian faces. Similarly, only Asian participants referenced the presence or absence of an eyelid crease in their descriptions. Although Asians noted “single eyelids” or “double eyelids” when describing both Asian and Caucasian faces, they did so with significantly greater frequency when viewing other Asian faces. These results suggest that Caucasians may realize that iris color is less useful for discriminating Asian faces (indicated by the lower proportion of iris color descriptors to Asian than to Caucasian faces) and that Asians may realize that the presence/absence of an eyelid crease is less useful for discriminating Caucasian faces (indicated by the lower proportion of eyelid crease descriptors to Caucasian than to Asian faces). Nevertheless, participants use cues appropriate for same-race face discrimination and apply them to cross-race face descriptions, consistent with predictions from the inappropriate cue utilization hypothesis. We discussed two possibilities for this behavior in the introduction: 1) using same-race appropriate cues to discriminate other-race faces precludes people from using more salient cues for cross-race discrimination and 2) people have not learned the most appropriate cues for cross-race discrimination. Results from Experiment 1 that participants are using same-race appropriate cues at a significantly lower rate to describe cross-race faces suggest that participants are not using same-race cues instead of cross-race cues, but rather they may not have learned the appropriate cross-race cues.
Feature Selection

Collapsing across participant race, the subjects in our sample mentioned race in their descriptions when viewing Asian test faces more than they mentioned race when viewing Caucasian faces. Asian participants also mentioned race more than Caucasians in their descriptions. Recall of Levin’s (2000) feature selection model and concept of the “race feature” that the model predicts both Asians and Caucasians would mention race with respect to other-race faces more than for same-race faces. Only the Caucasians in our sample followed this pattern. One explanation for this finding comes from literature surrounding rare and common events (see McKenzie, 2005 for a discussion on the rarity assumption). If you consider the general population demographics, Asians are in the minority and thus are more “rare” whereas Caucasians are more “common.” If people tend toward mentioning rare events, it follows that all participants in this experiment would mention the race of Asians faces more often than the race of Caucasian faces.

Experiment 1 Implications

The current literature surrounding the inappropriate cue utilization hypothesis fails to specify whether: A) people use those cues that are appropriate for discriminating same-race faces equally as often for discriminating other-race faces or B) people use those cues that are appropriate for discriminating same-race faces with a different frequency for discriminating other-race faces. Findings from the present study append the current literature by adding that although people appear to apply these less-appropriate or less-relevant cues to descriptions of other-race faces, they do so at a significantly lower rate than which they appropriately apply them to same-race faces. In other words, this may not be a matter of inappropriately using the wrong cues for cross-
race face discrimination, but rather failing to appropriately use the correct cues. This is an important distinction that the present study adds to our understanding of the own-race bias. Such findings suggest that people have some awareness that the cues used for same-race face recognition may not be as appropriate for other-race face recognition. The results also imply that although people may understand this at some level, they do not instead use those cues that are most relevant for other-race face discrimination.

When discussing the cues that participants used in the present study to describe faces, it is reasonable to think of iris color as being an appropriate cue to discriminate Caucasian faces, in part because Caucasian faces seem to vary along the dimension of iris color. Likewise, it is also sound to think of iris color as a less appropriate cue to discriminate Asian faces because Asians do not seem to vary greatly along this dimension. For the same reasons, we may consider the presence/absence of eyelid crease as a more appropriate cue to use for discriminating Asian faces but not for Caucasian faces. If in fact iris color is a more variable cue for Caucasian faces and eyelid crease is a more variable cue for Asian faces, do people generally understand that different races are not equally variable along these dimensions? Experiment 1 identified these cues as differentially significant to participants of each race, as evidenced through self-reports of important cues in a perceptual face description task. Experiment 2 tests whether these same features would affect performance in a memory task. If variability determines whether a feature is diagnostic, then we predict that participants will be sensitive to variations (i.e. manipulations between study and test) in a feature in accord with its expected range of variation for faces of a given race.
CHAPTER 3
Experiment 2

Introduction

Experiment 1 provided evidence that Asians and Caucasians refer to different facial features when describing faces. Further, they describe cues more variable for same-race faces more often when describing faces of the same race than faces of another race. Referring back to the concept of the ideal observer (e.g. Geisler, 1989), this model advocates that viewers should use the cues that are most variable, and thus most diagnostic, when recognizing faces. Applying this model to the results from Experiment 1 suggests that iris color is diagnostic for discriminating or recognizing Caucasian faces and that the presence/absence of an eyelid crease is diagnostic for discriminating Asian faces.

In general, people tend to look at the eye region of faces, in part because this area contains socially relevant information including social signals and affect indicators (Benuzzi, Pugnaghi, & Meletti, et al., 2007). Researchers have demonstrated that people look at the eye region not only to obtain this social information, but for purposes of face recognition as well. In a variety of recognition tasks, it has been demonstrated that subjects tend to look at the upper portion of the face more than the lower portion and pay disproportionate attention to the eyes (e.g., Shepherd, 1981; Althoff & Cohen, 1999). For example, Barton, Radcliffe, and Cherkasova, et al. (2006) used an eye-tracking paradigm and found a scanning preference for the eyes more than any other feature on the face. Subjects in this experiment also scanned the eye region more in novel faces than in famous faces.
Note that Experiment 1 provided evidence for the areas of the eye that different races describe when viewing faces but an explicit face description task does not necessarily provide data for where people are implicitly attending. Furthermore, it does not specify whether these features (whether they are implicitly or overtly perceived) play a role in face memory. We look to the types of cosmetic alterations that different races make to the eye region in order to provide further clues to which aspects of the eyes are attended to (and perhaps have more variance) by different races. We think of cosmetic alterations as changes aimed at increasing one’s beauty. What makes a face beautiful? Research suggests that average faces, rather than more distinctive or extreme faces, are rated as more attractive and beautiful (e.g. Langlous & Roggman, 1990; Rhodes & Tremewan, 1996). Our assumption is that people will only choose (on average) to alter their appearance within accepted norms, and that a change outside the range of normal variation would appear odd and perhaps reduce rather than enhance attractiveness.

Among Caucasians, for instance, there is a prevalence of altering iris color with colored contact lenses. Caucasians are born with a wide range of iris colors, including brown, blue, green, and hazel. Since the introduction of soft, colored contact lenses, Caucasians have been altering their eye colors within this range of natural variability. The Contact Lens Council reports that of the 85% of contact lens users who use soft contacts, 65% wear colored contact lenses (Contact Lens Council, 2000). See Experiment 1, Figure 2 for an example Caucasian eye with different colored contact lenses. Asians, on the other hand, are not born with a wide range of iris colors, but they are born with variance in the upper eyelid; some Asians have a crease in the upper eyelid while others do not. Among Asian societies, there exists a market for Asian
blepharoplasty, or double eyelid surgery. According to the American Academy of Facial Plastic and Reconstructive Surgery, 39% of plastic surgeries performed on Asian Americans in 2007 were blepharoplasties (AAFPRS, 2007). Such surgery creates a double eyelid by sewing a crease in the upper eyelid of Asians born with a single eyelid (an upper eyelid without a crease). See Experiment 1, Figure 3 for an example Asian eye before and after blepharoplasty surgery.

Interestingly, Asians and Caucasians make different types of cosmetic alterations to the eyes. This suggests firstly that there is a difference in eye features that the races generally attend to. It further suggests that these features have different ranges of variability for the races. Specifically, Caucasians display significant variance in iris color while Asians nearly always have a dark eye color. Asians, on the other hand, display significant variance in the presence or absence of an eyelid crease while Caucasians nearly always have this crease and thus double eyelids.

In Experiment 2a, we first question whether people are generally sensitive to those features that appear to have differential variability among races: 1) the natural variation in iris color for Caucasian faces and not for Asian faces and 2) the natural variation in presence/absence of an eyelid crease in Asian faces and not for Caucasian faces. We also question how these differential ranges of variability relate to face memory for Asian and Caucasian faces. If people are in fact sensitive to these natural variations in aspects of the eyes for different races, we predict the following: 1) varying eye color on Caucasian test faces will appear normal and difficult to detect, potentially resulting in increased false alarm rates, 2) varying eye color on Asian test faces will be easily recognized as incoherent, resulting in lower false alarm rates, 3) varying eyelid creases
on Asian test faces will appear normal, and participants will demonstrate higher false alarm rates, and 4) varying eyelid creases on Caucasian test faces will appear unusual, and participants will exhibit lower false alarm rates.

Experiment 2a uses a design that might focus participants’ attention on noticing variations in features outside the expected range of variability. In the second experiment (Experiment 2b), we follow up on Experiment 2a by only varying features within their expected ranges of variability for Asian and Caucasian faces. The inappropriate cue utilization hypothesis would predict that Caucasians attend more to iris color whereas Asians attend more to the presence/absence of an eyelid crease. If, for example, an Asian face is altered solely with respect to the eyelid crease, Asians would hypothetically notice this change more than would Caucasians. In a similar light, if a Caucasian face is altered with respect to iris color only, Caucasians would theoretically notice this change more than would Asians.

We evaluate these hypotheses in two experiments using a standard face recognition paradigm. Caucasian and Asian subjects view a set of target faces and then are asked to select the identical target faces from a mixed set of target and distracter faces. In this study, a portion of the distracter faces were created by altering the target faces with respect to race-appropriate (iris color in Caucasians and eyelid crease in Asians) and race-inappropriate cues (iris color in Asians and eyelid crease in Caucasians). We are primarily interested in the false alarm rates (reasons to be discussed in detail later) on completely new Asian and Caucasian faces as compared to false alarm rates on new Asian and Caucasian faces altered on the basis of iris color or eyelid crease. Research has demonstrated that false alarm rates can indicate the general nature of
information stored in memory (e.g. Shiffrin, Huber, & Marinelli, 1995; Roediger & McDermott, 1995), and we use false alarm rates to understand whether people are encoding changes in race-appropriate features into memory.

**Experiment 2a Method**

**Participants**

All participants ($N = 267$) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 29% male, 30% Caucasian, and 70% Asian.

**Stimuli**

The stimuli consisted of 80 pictures (225 x 311 pixels, gray scale): 40 Caucasian men and 40 Asian men. The original photographs were provided by Dr. Jim Tanaka\(^2\). Each of the 40 Asian faces had been placed inside an identical hairline with the same background and lighting; the same was true for the 40 Caucasian faces. All faces had neutral expressions and contained no facial hair or jewelry. Three forms of each stimulus face were used: 1) original face, 2) opposite eyelid, and 3) opposite iris. Of the “original” faces, half of the Asian test stimuli had single eyelids and half had double eyelids; similarly, half of the Caucasian test stimuli had light irises and half had dark irises. The “opposite eyelid” faces switched the style of eyelids on each face either from single eyelids to double eyelids or vice versa. Similarly, the “opposite iris” faces switched the darkness of the irises on each face either from light irises to dark irises or vice versa. Both the “opposite eyelid” and “opposite iris” faces were created by altering the original

---

\(^2\) Dr. Jim Tanaka, Professor, Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, British Columbia, Canada V8W 3P5. jtanaka@uvic.ca
pictures using Adobe Photoshop. See Experiment 2, Figure 1 for examples of the test stimuli. Participant responses were collected on the keyboard using the numeric keys. Stimuli presentation as well as participant response data were controlled and collected by an E-Prime script.

**Procedure**

The procedure used a traditional face recognition paradigm. Participants first viewed a set of 60 faces: 30 Caucasian men and 30 Asian men. Half of the Caucasian faces ($N = 15$) had light irises while the other half had dark irises, and half of the Asian faces had single eyelids while the other half had double eyelids. These faces were presented for three seconds each, in random order for each participant. Participants then completed a short distracter task where they saw old yearbook photos of famous celebrities and were asked to silently name each face (no response was required). They then viewed a set of 80 faces: 40 Caucasian men and 40 Asian men. Each face was presented for three seconds and then remained on the screen while participants indicated whether the face was old (they had seen it in the first set of faces) or new (they had NOT seen it in the first set of faces). Of the 80 faces in the test set: 20 (5 Asian with single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were truly old faces from the original study set; 20 (5 Asian with single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were new faces never seen before (which we refer to as the control condition); 20 (5 Asian with single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were faces from the original study set, except containing the opposite eyelid as the face seen during study; and 20 (5 Asian with
single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were faces from the original study set, except with the opposite iris as the face seen at study.

After completing this computer-based task, participants completed a brief survey (identical to that used in Experiment 1) of personal demographic information including zip codes of where participants were born and of where they have lived the longest in life.

Data Analysis

For each participant, we calculated two measures of recognition: 1) hits, the number of correct positive identifications (of those faces that were actually old) and 2) false alarms, the number of incorrect positive identifications in each of three conditions (those faces that were actually new, faces with opposite eyelids, and faces with opposite irises). We calculated the proportion of hits for each participant (of those faces that were actually old) in order to calculate \( d' \) (based on the proportion of hits on truly “old” faces and the proportion of false alarms on truly “new” faces). We report \( d' \) only to establish that participants demonstrate the ORB. “New” faces with swapped eyelids or swapped irises were not included in this analysis because this would not give us a baseline measure of whether participants were exhibiting the ORB.

We focus on false alarms in order to compare participants’ responses across the three conditions; there is only one type of target in relation to the three different types of distracters. Therefore, a comparison of false alarm rates is qualitatively the same as a comparison of unbiased sensitivity (\( d' \)). For this reason, we will not discuss results across conditions in terms of the signal detection measure of \( d' \). We are primarily
interested in the ratio of false alarms on faces with swapped features to false alarms on completely new faces. Therefore, we also calculated for each participant the difference in the proportion of false alarms between the opposite eyelid condition and the control condition as well as this difference between the opposite iris condition and the control condition.

**Experiment 2a Results**

We first used the proportions of hits and false alarms on the completely “old” and completely “new” faces to calculate $d'$ for each participant. We conducted a $2 \times 2$ ANOVA, using test face race (Asian, Caucasian) as a within-subjects independent variable, participant race (Asian, Caucasian) as a between-subjects independent variable, and $d'$ as the dependent variable. There were no significant main effects of participant race ($F(1,266) = 2.63, p > .05$) or test face race ($F(1,266) = 1.92, p > .05$) on $d'$. There was a significant interaction, however, between participant race and test face race ($F(1,266) = 30.07, p < .001$) on $d'$, indicating that participants in our sample were demonstrating the own-race bias with respect to $d'$ on original test faces that were not altered (either on iris color or eyelid crease). Specifically, Asian participants showed a higher $d'$ for Asian faces ($M = .64$) than for Caucasian faces ($M = .40$) while Caucasian participants conversely showed a higher $d'$ for Caucasian faces ($M = .62$) than for Asian faces ($M = .23$). All participants showed greater discrimination ability for same-race faces than for other-race faces.

Because hit rates do not inform as to participants’ performance on those new test faces with swapped eyelids or swapped iris colors (as discussed before), we examined the effects of participant race and test face race on false alarm rates across the three test face
conditions. We conducted a 3 x 2 x 2 ANOVA, using condition (control, opposite eyelid, opposite iris) and test face race (Asian, Caucasian) as within-subjects independent variables, participant race (Asian, Caucasian) as a between-subjects independent variable, and the proportion of false alarms as the dependent variable. Firstly, there was a significant interaction between condition and test face race. Participants false alarmed significantly more on Asian faces \((M = .544)\) than on Caucasian faces \((M = .489)\) in the opposite eyelid condition and false alarmed more on Caucasian faces \((M = .593)\) than on Asian faces \((M = .276)\) in the opposite iris condition, \(F(2,265) = 114.14, p < .001\) (see Experiment 2, Graph 1).

We also found main effects of condition, test face race, as well as participant race on the number of false alarms. Participants made significantly more false alarms in the opposite eyelid condition \((M = .514)\) than in the opposite iris condition \((M = .426)\) or in the control condition of entirely new faces \((M = .441)\), \(F(2,265) = 26.49, p < .001\). Also, Asian participants \((M = .472)\) made more false alarms than Caucasian participants \((M = .449)\), \(F(1,266) = 4.98, p < .05\), and all participants made more false alarms when viewing Caucasian faces \((M = .504)\) than when viewing Asian faces \((M = .417)\), \(F(1,266) = 66.74, p < .001\). There was also a significant interaction between participant race and test face race. When viewing Asian faces, Caucasian participants \((M = .432)\) exhibited a greater false alarm rate than Asian participants \((M = .402)\); likewise, when viewing Caucasian faces, Asian participants \((M = .542)\) false alarmed more often than Caucasian participants \((M = .465)\), \(F(1,266) = 25.27, p < .001\). This interaction between participant race and test face race indicates that participants in our sample were demonstrating the own-race bias.
However, despite this finding of an ORB in recognition, there was no significant 3-way interaction between participant race, test face race, and condition.

We calculated the difference scores for each participant (difference between the number of false alarms in the opposite eyelid condition and control as well as difference between the number of false alarms in the opposite iris condition and control) and conducted a 2 x 2 x 2 ANOVA using condition (opposite eyelid, opposite iris) and test face race (Asian, Caucasian) as within-subjects independent variables, participant race (Asian, Caucasian) as a between-subjects independent variable, and the difference in d' as the dependent variable. Again, there was a significant interaction between condition and test face race. There were significantly more false alarms on Asian faces (M = .13) than on Caucasian faces (M = .01) in the opposite eyelid condition, and participants false alarmed more often on Caucasian faces (M = .12) than on Asian faces (M = -.14) in the opposite iris condition, F(1,266) = 122.47, p < .001 (see Experiment 2, Graph 2). We also found main effects of test face race (F(1,266) = 30.28, p < .001) and condition (F(1,266) = 28.42, p < .001) on the proportion of false alarms. There was a higher average increase in false alarm rate in the opposite eyelid condition compared to the control (M = .07) than in the opposite iris condition compared to the control (M = -.02). These findings are consistent with the results we obtained from the initial analysis.

Interracial Contact

We investigated whether one’s self-reported level of interracial contact influenced the proportion of false alarms to both Asian and Caucasian faces. If an Asian participant lived the longest in an area where the proportion of Caucasians surpassed the proportion of Asians, this participant was coded as a high-contact participant (56%). Otherwise, the
participant was considered low-contact (44%). Because only one Caucasian in this sample reported living the longest in an area where the proportion of Asians surpassed the proportion of Caucasians, we focus our interracial contact analyses solely on the Asian participants.

We performed a 3 x 2 x 2 ANOVA with level of contact (high-contact Asians, low-contact Asians) as a between-subjects independent variable, test face race (Asian, Caucasian) and condition (control, “opposite eyelid,” “opposite iris”) as within-subjects independent variables, and false alarm rate as the dependent variable. We did not find a significant main effect of the level of interracial contact on false alarm rate for the Asian participants in the sample, $F(1,186) = 2.12, p > .05$. Similarly, we did not find an interaction between the level of interracial contact and either the race of the test face ($F(1,186) = .20, p > .05$) or the condition ($F(2,185) = .04, p > .05$) on the false alarm rate. Because there was no main effect of condition, we collapsed across condition and performed a 2 x 2 ANOVA with level of contact (high-contact Asians, low-contact Asians) as a between-subjects independent variable, test face race (Asian, Caucasian) as a within-subjects independent variable, and false alarm rate as the dependent variable. Again, there was no main effect of level of contact on false alarm rate ($F(1,186) = 1.80, p > .05$), nor was there an interaction between level of interracial contact and test face race ($F(1,186) = .17, p > .05$). High-contact Asians were equally likely to false alarm on Asian and Caucasian faces ($M = .41, M = .55$) as were low-contact Caucasians ($M = .39, M = .53$). The contact hypothesis predicts that Asians with increased interracial contact with Caucasians would demonstrate more accurate recognition of Caucasian faces than Asians with less contact, and that this contact should not mediate recognition accuracy of
same-race Asian faces. Our results, however, reflect no impact of interracial contact alone on subsequent recognition accuracy of Asian or Caucasian faces.

**Experiment 2a Discussion**

In this first experiment, both races of participants demonstrated the own-race bias, with Asians making more false alarms on Caucasian faces than on Asian faces and Caucasians false alarming more often on Asian than Caucasian faces. Further, Caucasians showed a higher $d'$ for Caucasian than for Asian test faces and Asians showed a higher $d'$ for Asian than for Caucasian test faces. We predicted that people in general would be more sensitive to changes in iris color on Asian faces and to changes in eyelid creases on Caucasian faces because such changes would be outside the expected range of variation. Therefore, our prediction was for higher false alarms to eyelid crease changes on Asian faces and higher false alarms to iris color changes on Caucasian faces because such changes are within the expected ranges of variation (and thus less likely to be noticed) for each race. Referencing Graph 2.2, we can see that participants did in fact exhibit higher false alarm rates to Asian faces with swapped eyelid creases and to Caucasian faces with swapped irises. Further, participants demonstrated a significant decrease in false alarms to Asian faces with swapped iris colors. This is consistent with our prediction that participants would be sensitive to the fact that there is low variability in Asian iris color and in presence/absence of Caucasian eyelid creases; when changes were made outside the expected range of variability (particularly for Asian test faces), participants noticed and subsequently exhibited a lower false alarm rate. However, when changes were made to faces within the expected range of variability, participants did not notice and exhibited higher false alarm rates.
We also tested whether Caucasian participants would be more sensitive to changes in iris color on Caucasian and Asian test faces than would Asian participants and whether Asian participants would be more sensitive to changes in the presence/absence of eyelid creases. Our results indicate that both Caucasians and Asians are equally sensitive to changes in iris color, particularly more so for Asian faces than for Caucasian faces. Similarly, both races are insensitive to changes in eyelid creases, particularly for Asian faces (see Experiment 2, Graph 2). It may be best to conceptualize these findings in the following manner: people are sensitive to changes in features when these changes alter the features outside the expected range of variability. Because iris color has a naturally restricted range on Asian faces, altering this color on Asian faces alerted all participants to changes outside the typical range. Likewise, Caucasian faces have a more restricted range of eyelid crease variability than do Asians, and thus participants false alarmed more on eyelid crease changes to Asian faces.

Results from Experiment 2a imply that people are generally sensitive to changes made outside an expected range of variability for a given feature, but are not as sensitive when these changes fall within the expected range. It is possible that participants only attended to the obvious changes in features outside the expected ranges of variability for iris color (particularly for Asian faces) and presence/absence of an eyelid crease (particularly for Caucasian faces) and thus did not bother attending to possible, more subtle changes within the expected range. Essentially, Experiment 2a may have placed participants in a situation of looking for test faces that looked distinctly odd. Such a response strategy would explain the large decreases in false alarms with changes outside the ranges of expected variation. Therefore, in order to rule out such a response strategy,
Experiment 2b only uses changes within the ranges of expected variation so as not to bring overt notice to the changes.

Experiment 2b compares Caucasian and Asian participants’ recognition of Caucasian and Asian faces when Caucasian (but not Asian) faces are altered with respect to iris color and Asian (but not Caucasian) faces are altered on the presence/absence of an eyelid crease. Furthermore, unlike Experiment 2a, we use equal numbers of altered face images at study and at test. If participants noticed the altering of the faces through image artifacts, then they may have used this information to know that a test face was a distracter rather than a target considering that more altered faces appeared as distracters on the test list and few appeared on the study list. By including an equal mix of altered faces at study and at test, this establishes that noticing an altered face cannot be used as a cue to know whether the test face is a target versus a distracter.

In Experiment 2b, we ask whether participants are differentially sensitive to changes in features that appear to be more appropriate (because they are more variable) for one race than another when these changes are made within the expected range of variability for that race. We predict that Caucasians will be more sensitive (exhibit fewer false alarms) than Asians to Caucasian faces altered on the basis of iris color, and Asians will be more sensitive than Caucasians to Asian faces altered on the presence/absence of an eyelid crease.

Experiment 2b Method

Participants
All participants (N = 91) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 25% male, 33% Caucasian, and 66% Asian.

**Stimuli**

The stimuli were identical to the stimuli used in the first study. Stimuli presentation as well as participant response data were again controlled and collected by an E-Prime script.

**Procedure**

The procedure was extremely similar to the face recognition paradigm used in Experiment 2a. Participants first viewed a set of 40 faces: 20 Caucasian men and 20 Asian men. Half of the Caucasian faces (N = 10) had light irises while the other half had dark irises, and half of the Asian faces had single eyelids while the other half had double eyelids. These faces were presented for three seconds each, in random order for each participant. Participants then completed a short distracter task where they saw old yearbook photos of famous celebrities and were asked to silently name each face (no response was required). They then viewed a set of 60 faces: 30 Caucasian men and 30 Asian men. Each face was presented for three seconds and then remained on the screen while participants indicated whether the face was old (they had seen it in the first set of faces) or new (they had NOT seen it in the first set of faces). Of the 60 faces in the test set: 20 (5 Asian with single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were truly old faces from the original study set; 20 (5 Asian with single eyelids, 5 Asian with double eyelids, 5 Caucasian with light irises, 5 Caucasians with dark irises) were new faces never seen before (which we refer to as the
control condition); 10 were Asian faces from the original study set (5 with single eyelids, 5 with double eyelids), except containing the opposite eyelid as the face seen during study; and 10 were Caucasian faces from the original study set (5 with light irises, 5 with dark irises), except with the opposite iris as the face seen at study.

Data Analysis

For each participant, we calculated six measures of recognition: 1) the number of hits on Asian faces, 2) the number of hits on Caucasian faces, 3) the number of false alarms on Asian faces, 4) the number of false alarms on Caucasian faces, 5) the number of false alarms on Asian faces with swapped eyelids, and 6) the number of false alarms on Caucasian faces with swapped iris colors. We calculated the proportion of correct identifications for each participant (of those faces that were actually old in the control condition) in order to calculate $d'$ (based on the proportion of hits on truly “old” faces and the proportion of false alarms to truly “new” faces). We again report $d'$ only to establish that participants demonstrate the ORB. “New” faces with swapped eyelids or swapped irises were not included in this analysis because this would not give us a baseline measure of whether participants were showing the ORB or not. We focus on false alarms in order to compare participants’ responses across conditions; correct identification of an old face (a hit) is not possible in the opposite eyelid and opposite iris conditions, so we cannot compare hit rates across the three conditions.

Experiment 2b Results

We first used the proportion of hits and false alarms on the completely “old” and completely “new” faces to calculate $d'$ for each participant. We conducted a 2 x 2 ANOVA with participant race (Asian, Caucasian) as a between subjects independent
variable, test face race (Asian, Caucasian) as a within subjects independent variable, and
d' as the dependent variable. There were no significant main effects of participant race
\( F(1,90) = .66, p > .05 \) or test face race \( F(1,90) = .16, p > .05 \) on d'. There was a
significant interaction, however, between participant race and test face race \( F(1,90) = 
4.89, p < .05 \) on d', indicating that participants in our sample were demonstrating the
own-race bias with respect to d' on original test faces that were not altered (either on iris
color or eyelid crease). Specifically, Asian participants showed a higher d' for Asian
faces (M = 1.10) than for Caucasian faces (M = .82) while Caucasian participants
conversely showed a higher d' for Caucasian faces (M = .97) than for Asian faces (M = 
.78).

Because hit rates do not inform as to participants’ performance on those new test
faces with swapped eyelids or swapped iris colors (as discussed earlier), we examined the
effects of participant race and test face race on false alarm rates across the four test face
conditions. We conducted a 2 x 4 ANOVA with participant race (Asian, Caucasian), and
condition of test face (Asian, Caucasian, Asian with swapped eyelid creases, Caucasian
with swapped irises) as independent variables and the proportion of false alarms as the
dependent variable. There was a significant main effect of test face condition such that
all participants demonstrated a significantly greater false alarm rate to the Asian faces
with swapped eyelid creases (M = .55) and the Caucasian faces with swapped irises (M = 
.57) than to the Asian faces with no alterations (M = .29) or Caucasian faces with no
alterations (M = .32), \( F(3,88) = 45.77, p < .001 \) (see Experiment 2, Graph 3).

There were no significant differences between Asian (M = .57) and Caucasian (M 
= .54) participants on the proportion of false alarms to Asian faces with swapped eyelid
creases, $t(90) = 0.66$, $p > .05$, nor were there any significant differences between Asian ($M = .58$) and Caucasian ($M = .55$) participants with respect to proportion of false alarms on Caucasian faces with swapped irises, $t(90) = 0.60$, $p > .05$ (also, see Experiment 2, Graph 3).

**Interracial Contact**

We again investigated whether one’s self-reported level of interracial contact affected the proportion of false alarms to both Asian and Caucasian faces. If an Asian participant lived the longest in an area where the proportion of Caucasians surpassed the proportion of Asians, this participant was coded as a high-contact participant (64%). Otherwise, the participant was considered low-contact (36%). Because no Caucasians in this sample reported living the longest in an area where the ratio of Asians surpassed that of Caucasians, we again focus our interracial contact analyses solely on the Asian participants.

We performed a $4 \times 2$ ANOVA with level of contact (high-contact Asians, low-contact Asians) as a between-subjects independent variable, condition (Asian faces, Caucasian faces, opposite eyelid Asian faces, opposite iris Caucasian faces) as a within-subjects independent variable, and false alarm rate as the dependent variable. We did not find a significant main effect of the level of interracial contact on false alarm rate for the Asian participants in the sample, $F(1,59) = .10$, $p > .05$. High-contact Asians false alarmed equally often as low-contact Asians. Similarly, we did not find an interaction between the level of interracial contact and condition on the false alarm rate, $(F(3,57) = .13, p > .05)$. 
Because there was no main effect of condition on false alarm rate for high- and low-contact Asians, we collapsed conditions across test face race to create two conditions: 1) Asian faces and opposite eyelid Asian faces and 2) Caucasian faces and opposite iris Caucasian faces. We then conducted another 2 x 2 ANOVA with level of contact (high-contact Asians, low-contact Asians) as a between-subjects independent variable, test face race (Asian, Caucasian) as a within subjects independent variable, and false alarm rate as the dependent variable. We found a significant main effect of face race on false alarm rate, replicating the finding that Asians false alarmed more often on Caucasian faces ($M = .49$) than on Asian faces ($M = .35$), $F(1,59) = 19.85, p < .001$. Of note, there was a significant interaction between level of interracial contact and test face race on false alarm rate such that the difference in false alarm rates on Asian ($M = .28$) and Caucasian faces ($M = .51$) was significantly larger for the low-contact Asians than was the difference in false alarm rates on Asian ($M = .41$) and Caucasian faces ($M = .47$) for high-contact Asians, $F(1,59) = 6.92, p < .01$. The low-contact Asians false alarmed less on Asian faces and more on Caucasian faces than did the high-contact Asians, indicating that interracial contact is indeed a moderator of cross-race recognition accuracy.

Experiment 2b Discussion

In this second study, we manipulated race-appropriate cues within their expected ranges of variability and found that participants were generally insensitive to these changes. We expected that Caucasian participants, in particular, would have fewer false alarms than Asians to Caucasian faces altered on the basis of iris color (from light irises to dark or vice versa). In fact, both races of participants were equally insensitive to these
changes in eye color. Similarly, we expected that Asian participants would have fewer false alarms than Caucasians to Asian faces with swapped eyelid creases (from single eyelids to double or vice versa). Again, all participants appeared equally insensitive to these eyelid crease changes.

Findings that participants had increased false alarms on faces that were altered only with respect to the specified race-appropriate cue suggests that perhaps one short study period is not enough for participants to differentially encode the appropriate features in memory. It may also suggest that while these cues are important for perception (as found in Experiment 1), they might not be easily included in the information stored in memory, particularly for novel faces only seen briefly. Thus, these cues may only affect memory performance if they are perceptually noticed at test as being odd (outside the range of expected variability) as results from Experiment 2a indicate. Essentially, these cues may be noticed in perception rather than stored in memory. There is evidence in the literature that the visual icon includes raw feature information (Sperling, 1960) whereas long-term memory tends to encode items semantically in terms of the general idea (Craik & Tulving, 1975). By analogy, it follows that facial features may be important in perceptual and descriptive analyses while holistic or configural information may be more important for face memory. In Experiment 2, although iris color may be a race-appropriate cue for Caucasian faces and the presence or absence of an eyelid crease may be race-appropriate for Asians, one short study task is insufficient for participants (even individuals who are more expert at using the appropriate cues for same-race face discrimination) to encode the appropriate information into memory.
Interracial Contact

There was a significant interaction between level of interracial contact (for Asian participants) and test face race on the proportion of false alarms such that Asian participants with less interracial contact false alarmed more on Caucasian faces and less on Asian faces than did those Asian participants with more interracial contact. This finding is consistent with the literature indicating that interracial contact is a clear moderator for ORB expression (e.g. Chance et al., 1975; Brigham, Maass, Snyder, & Spaulding, 1982; Brigham & Malpass, 1985; Meissner & Brigham, 2001). Though this experiment did not attempt to establish a relationship between interracial contact and learning appropriate cues, it nonetheless did demonstrate independently that certain cues are differentially relevant to different races and that differential levels of interracial contact influence recognition ability for other-race faces.

Interestingly, in Experiment 2a, there was no effect of interracial contact on false alarm rate, but in Experiment 2b, low-contact Asians false alarmed significantly more on Caucasian faces than did high-contact Asians. We explain this disparity by looking at the stimuli used in the two experiments. While Experiment 2a included faces that had features outside the natural range of variability for a given race, Experiment 2b did not. The test faces in Experiment 2b are more representative of faces that would naturally be seen in the real world, and thus perhaps a more valid and accurate gauge on which to measure the effects of interracial contact.

Summary of Experiment 2

Findings from Experiments 2a and 2b provide us with a more complete picture of our general ability to detect changes in features both within and outside of the expected
range of variability for a given feature. We examined two features that may be differentially appropriate to discriminate between faces of different races. Asian faces vary along the dimension of eyelid crease more than do Caucasian faces, and Caucasian faces vary more along the dimension of iris color. An ideal observer (e.g. Geisler, 1989) would use these race-appropriate cues to discriminate between faces because of their differential variability for Asians and Caucasians. Experiment 2a supplied evidence that people are generally sensitive to changes in these features when the changes appear outside of the natural range of variability. Furthermore, this sensitivity is not constrained to individuals of the same race for which the cue is most appropriate.

The fact that both races appear equally sensitive to changes in features outside the normal range even for cross-race faces may be a byproduct of our manipulation. We cosmetically altered these faces such that participants in Experiment 2a were able to detect that something about a face was amiss. We cannot determine whether participants understood the exact nature of the changes to the faces, only that participants detected that something was noticeably different about the faces. When we constrained the changes in Experiment 2b to alter features only within their normal ranges of variability for a given race, participants were no longer sensitive to these changes.

We can see that people express some level of awareness that the different aspects of the eyes vary more for some races than for others. Further, although both Asians and Caucasians demonstrated the own-race bias, they did not differentially notice or use changes in race-appropriate cues for memory of the faces. Experiments 2a and 2b were memory tasks demonstrating no interaction between participant race and use of race-appropriate cues for face memory. What this study does not tell us, however, is whether
different race participants perceptually attend to race-appropriate cues for either same-race or cross-race faces when this cue is within the normal range of variability.

Experiment 1 suggested that these cues are important to perception based on self-report. However, self-report may place an emphasis on specific, reportable features that one can put into words rather than holistic face information, which may be more difficult to explain. Therefore, Experiment 3 is aimed toward determining those features involved in perception through similarity ratings, which allows for a more Gestalt-like judgment without overtly referring to specific features.

We also must be careful not to think of cues as being dichotomously appropriate or inappropriate, but rather on a continuum of more or less appropriate. This raises the question, how do we determine the most appropriate cues that people use when discriminating faces? Experiments 2a and 2b only allowed us to infer that iris color may be more diagnostic for Caucasian faces and that presence/absence of an eyelid crease may be more diagnostic for Asian faces, but the results do not tell us the degree to which these cues are more appropriate for one race over another and how important they are for discriminating faces. Further, Experiment 1 only allowed us to gather information on the variety of cues that people attend to and spontaneously generate when describing faces. This particular measure does not allow us to determine those cues and features that people actually use most often opposed to those cues they report to be most important or useful when discriminating between faces; it is precisely this unanswered question that provides the motivation for Experiment 3.
CHAPTER 4

Experiment 3

Introduction

Experiment 2 demonstrated that there are cues that vary more for some races than others and as a result may be more diagnostic for discriminating between faces of that race. In particular, iris color varies more for Caucasian than Asian faces while the presence/absence of an eyelid crease varies more for Asian than Caucasian faces. Further, Experiment 1 showed that what we deem to be cues “appropriate” for discriminating between faces of a given race (in part because these features have high variability within the race) are in fact used more often by members of this race when describing faces. Moreover, members of this race use these own-race appropriate cues more often for describing same-race faces than other-race faces. In Experiment 3, we use global similarity ratings to extract not only those features that participants use to describe faces (if they are even used in such Gestalt-like similarity ratings) but the other features that people actually use when discriminating faces. Furthermore, these global ratings allow us to place these features into a broader context as compared to highly salient facial features such as race and gender.

Multidimensional scaling (MDS) may unveil the most important psychological (not ecological) dimensions along which same-race and other-race faces are perceived. For a full discussion on multidimensional scaling, see Kruskal & Wish (1986). In brief, MDS is a technique that starts with the average similarity ratings for all possible pairs of objects in a stimulus set (in this case, the stimuli are faces). MDS then allows for visual representation of these ratings for all pairs that exist in N-dimensional space. Each point
in this visual configuration corresponds to one of the objects in the stimulus set. The closer two points, the more similar they were rated. MDS allows researchers to determine the “hidden structure” (Kruskal & Wish, 1986, p. 7) in the similarity ratings between the pairs. For a stimulus set of faces, this structure is a set of dimensions such as race, hair color, gender, or face shape based on participants’ similarity ratings of all face pairs. These dimensions are the most important cues that participants use to rate the similarity between the faces. It is currently unknown whether: 1) different race individuals use different dimensions to discriminate faces or 2) people in general use fewer dimensions, or different dimensions altogether, to discriminate other-race faces than to discriminate own-race faces. Further, it is likely, based on the findings from Experiments 1 and 2 that individuals of different races use different dimensions when comparing or discriminating between faces.

Other researchers have used MDS to uncover what features are used in face perception. In particular, the literature suggests that MDS reveals race as one of the primary dimensions depicted in a face-space (e.g. Busey, 1998) such that other-race faces are represented in a smaller, more dense grouping (indicating these faces were rated as highly similar) and same-race faces are represented in a larger, more disperse cluster (indicating that these faces were rated as highly dissimilar) (Levin, 1996). This is based in part on Shepherd and Deregowski’s (1981) task asking Whites and Blacks to sort trios of faces on the basis of different cues (two faces being most similar and one face being more dissimilar on X cue) and then rate each face on the cue. Multidimensional scaling analysis was then applied to participants’ ratings and they found that participants made judgments on faces using different cues for White than for Black faces. Replicating the
study with a homogenous set of Black test faces, Black participants discriminated faces using a greater variety of cues than did White participants. Valentine (1991) used Shepherd and Deregowski’s (1981) results to explain the clustering of same- and cross-race faces in a multidimensional face space; same-race faces will cluster around one central region and other-race faces will cluster around a different central region. The important difference, Valentine argues, is that cross-race faces are clustered more densely than are same-race faces. Chiroro & Valentine (1995) have used this multidimensional face space concept as an explanation for the own-race bias in that the dense cluster of other-race faces lies in a sparse region of the face space compared to the area occupied by same-race faces. Consequently, these other-race faces may be perceptually (but not necessarily physically) more similar, thus increasing the difficulty of other-race discrimination.

To date, researchers have only looked at MDS solutions (generated from pairwise similarity ratings on a set of stimuli) based on ratings from one homogenous group of participants; researchers have not looked at how MDS solutions (for the same set of stimuli) vary for different participant races. Busey (1998), for example, asked participants to provide similarity ratings for a set of bald, male, primarily Caucasian photos. The resulting MDS solution revealed the following six dimensions in order of quantity of variance accounted for: age, race, facial adiposity (pudginess), facial hair, aspect ratio of head (tall and thin versus short and squat), and color of facial hair. Similarly, Johnston et al. (1997) asked a small number of participants (N = 12) to rate the similarity between adult male (race unspecified) photos. Researchers identified the dimensions of face width, forehead size, and perceived age. As demonstrated by
Johnston and colleagues, researchers cannot always guarantee that all emerging dimensions will be positively identified. Busey (1998) argues, however, that the more dimensions that are identifiable, the more valid the face space model.

The literature presently lacks a complete cross-race multidimensional scaling investigation using similarity ratings with two different race participants viewing a set of stimuli containing faces from both participant races. This is the appropriate data needed to gain a full understanding of the dimensions that different races use when evaluating faces. In this experiment, we conduct a multidimensional scaling analysis with both Asian and Caucasian participants. Participants rate all possible pairs of a set of pictures containing both Asian and Caucasian faces. We conduct two separate analyses for the Asian and Caucasian participants in order to understand the dimensions that both races use for face discrimination and the differential importance (if any) that people place on different cues. In this study, we are concerned with uncovering the internal (inside the hair line) dimensions of a face that people when judging the similarity between faces. To achieve this, all test faces shown to participants have an identical hair line; this manipulation prevents participants from relying upon external dimensions such as hair color, style, and texture and forces them to focus on the actual features of a face.

Based on the findings from Experiment 1 that Caucasians refer to iris color more often than do Asians and Asians refer to the presence/absence of an eyelid crease more often than do Caucasians, we hypothesize that if iris color emerges as a dimension, it will do so either: 1) only in the Caucasian MDS solution or 2) as a stronger dimension in the Caucasian than in the Asian MDS solution. Recall that in Experiment 1, both Caucasians and Asians mentioned iris color in their face descriptions, but Caucasians did so with
significantly greater frequency than did Asians. Further, we expect that if eyelid crease emerges as a dimension, it will only appear in the Asian MDS solution. This prediction is based on the finding in Experiment 1 that only Asian subjects mentioned single and double eyelids in their descriptions. We cannot predict the exact dimensions that will emerge in this multidimensional scaling research because MDS is an exploratory investigation. For the same reason, we cannot predict the precise order in which the various dimensions will appear, nor can we anticipate that we will be able to identify every dimension (e.g. Busey, 1998).

Further predictions come from the contact hypothesis such that individuals with more interracial contact should have MDS solutions that more closely (than individuals with less interracial contact) resemble the other-race MDS solution. Specifically, Caucasians high on interracial contact with Asians would more likely show eyelid crease as a dimension in their MDS solution than would Caucasians low on interracial contact. Similarly, Asians high on interracial contact with Caucasians would be more likely to show iris color as a dimension (or as a more important dimension) than would Asians low on interracial contact. We assess interracial contact with a questionnaire that not only gathers information on participants’ self-reported levels of interracial contact or on the demographic data of the area where participants are currently living at the time of the investigation, as have been common methods reported in the literature (e.g. Malpass & Kravitz, 1969; Chance et al., 1975; Brigham & Barkowitz, 1978), but also with demographic data about the area where participants have spent the most time in their lives. This is a novel addition to the methods used for determining the quantity/quality of participants’ interracial contact.
Method

Participants

All participants (N = 162) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 33% male, 33% Caucasian, and 67% Asian.

Stimuli

The stimuli consisted of 40 pictures (300 x 400 pixels, gray scale): 10 Caucasian men, 10 Caucasian women, 10 Chinese men, and 10 Chinese women. All of the Asian faces had dark irises, but were mixed on the presence/absence of eyelid crease. Conversely, all the Caucasian faces had double eyelids, but were mixed on the darkness of iris color. The original photographs were provided by Dr. Kang Lee and taken under similar lighting and background. These are the same photographs that were used in Experiment 1. The difference between the pictures in this experiment and those in Experiment 1 lies in the hairline of the faces. In Experiment 1, all faces had their original hair. For this study, each face was placed inside a neutral hairline that was created by morphing one Caucasian male, one Caucasian female, one Chinese male, and one Chinese female (see Experiment 3, Figure 1 for examples of the test stimuli). With this design, we were able to eliminate the possibility that subjects relied on features external to the face when judging the similarity between any two given faces. In particular, we wanted to eliminate hair as a marker or cue for race and gender. All faces had neutral expressions and contained no facial hair or jewelry.

The 40 faces produced 780 unique face pairs, which were presented to subjects simultaneously side by side in random position (left/right) and order. Twenty-eight of the
126 participants viewed all 780 pairs of faces. All other participants viewed only 390 pairs of faces, such that two participants [independently] viewed the two halves from each random order of face pairs. Participant responses were collected on the keyboard using the numeric keys 1-7. Stimuli presentation as well as participant response data were controlled and collected by an E-Prime script.

**Procedure**

Participants rated the similarity of 390 face pairs on a 7-point scale (1 = very different, 7 = very similar). Each pair of faces remained on the screen for 3 seconds, after which time participants indicated the similarity rating for the previously viewed pair. Participants received a short break after every 50 pairs of faces in order to minimize participant fatigue. Each of the 780 face pairs was rated by an average of 35 Caucasian participants as well as 66 Asian participants.

After completing the similarity rating task, participants completed a survey aimed at assessing their quantity and quality of prior contact with persons of the “other race” (either Caucasian or Asian). See Experiment 3, Figure 2 for a complete copy of the questionnaire. The questionnaire was built upon the foundations of the interracial contact questionnaire created by Malpass and Kravitz (1969). The questionnaire probed the quantity and quality of interracial contact via: 1) 24 questions regarding the proportion of both same- and other-race individuals encountered in different roles throughout the participants’ lives, 2) demographics and frequency of contact with participants’ three closest friends, and 3) personal demographic information including zip codes of where participants were born and of where they have lived the longest in life.

**Results**
There were no significant differences in the pattern of responses between those participants who viewed the entire set of 780 face pairs and those participants who viewed half of the face pairs. Therefore, the reported results are collapsed across all participants.

Similarity ratings

Caucasian participants (M = 3.22) rated the face pairs to be more similar overall than did the Asian participants (M = 2.86). This pattern is significant for face pairs of 1) both Asians (t(161) = 24.66, p < .001), 2) both Caucasians (t(161) = 9.61, p < .001), and 3) one face of each race (t(161) = 27.03, p < .001), with Caucasians (M = 4.14, 3.60, 2.60) judging the faces as appearing more similar to one another than Asians (M = 3.56, 3.39, 2.27). Notably, both Asian and Caucasian participants followed the same pattern in assessing pairs of Asian faces as being more similar to one another on average than pairs of Caucasian faces. Additionally, both groups judged pairs with one face from each race as being significantly less similar than pairs of faces from either single race, F(2,160) = 26.19, p < .05.

Environmental analysis of similarity

In examining whether one race was in fact more physically homogenous than the other, as suggested above, twelve separate measurements as well as twelve ratios calculated from these actual measurements were obtained for each of the 40 faces. The measurements included the following distances: right eye to right ear; left eye to left ear; right ear to right side of mouth; left ear to left side of mouth; between the eyes; mouth width; mouth length; chin length; bottom of nose to top of lips; nose width; nose length; and forehead length. The ratios between these measurements included: nose width to
nose length; mouth width to mouth length; nose length to mouth length; nose width to
mouth width; nose width to distance between nose and mouth; nose length to distance
between nose and mouth; forehead length to chin length; chin length to mouth length;
mouth length to distance between nose and mouth; chin length to distance between nose
and mouth; nose length to forehead length; and length of top of the face to length of
bottom half of the face (divided at the base of the nose).

The physical measurements indicate that neither the sample of Asian nor
Caucasian faces used in the present study is truly more homogenous than the other. In
fact, although participants indicated that the Asians appeared more similar to one another
than the Caucasians, the Asian faces had a significantly greater variance on one of the
twelve physical measurements (chin length), $F(1,19) = 2.24$, $p < .05$. On the other eleven
physical measurements and on the twelve ratios calculated between these measurements,
neither the Asian faces nor the Caucasian faces were significantly more variable. This
indicates that the Asian faces actually were no more physically similar to one another
than were the Caucasian faces.

Evaluating the MDS solutions

The confusion matrix of the similarity ratings for all 780 pairs of faces was
analyzed using the ALSCAL scaling procedure. Two separate ALSCAL analyses were
completed for Asian and Caucasian participants. The resulting six-dimensional MDS
solution from the Asian participants’ data had an $r^2$ value of .945 and a stress value of
.069. In MDS, stress is a measure of the degree of correspondence between the ideal and
actual distances among the points (Kruskall & Wish, 1978). It is measured such that the
lower the stress, the greater the correspondence between ideal and actual distances, and
thus the more reliable the MDS solution. Similar to the Asian participants, the six-dimensional MDS solution resulting from the Caucasian participants’ data had an $r^2$ value of .939 and a stress value of .074. Both MDS algorithms accounted for a large share of the variability within the similarity ratings. Further, the stress values were significantly low for both solutions to indicate a reliable algorithm (e.g. Kruskall & Wish, 1978; Sturrock & Rocha 2000).

As noted by Busey (1998), one test for the strength of an MDS model is whether the resulting dimensions are interpretable. Refer to Figures 3.3 and 3.4 for the distribution of faces along the six dimensions for the MDS solutions from Asian and Caucasian participants, respectively. The six dimensions from each model represent those dimensions along which the participants relied most heavily when rating the similarity between the pairs of faces in the study. Experiment 3, Table 1 presents correlations between the six dimensions that emerged for Asian and Caucasian MDS models. Dimension 1 was visually and statistically identifiable for both Caucasian and Asian participants as that of race. For Caucasians, the race of the test face correlated strongly with Dimension 1 ($r(38) = .966, p < .01$), and for Asians, the race of the test face also correlated strongly with Dimension 1 ($r(38) = .963, p < .01$). This indicates that race is the factor most heavily relied upon by Asians and Caucasians alike when comparing faces. Visually and statistically, Dimension 2 emerged for both groups as that of gender. For Caucasians, the gender of the test face correlated strongly with Dimension 2 ($r(38) = .883, p < .01$), and for Asians, the gender of the test face also correlated strongly with Dimension 2 ($r(38) = .887, p < .01$). Dimensions 1, 2, and 3 correlate directly between Asian and Caucasian participants, signifying that those dimensions represent identical
factors for both races of participants in the study. Specifically, Dimension 1 from the Asian MDS solution correlates with Dimension 1 from the Caucasian MDS solution ($r(38) = .996, p < .01$), Dimension 2 from the Asian MDS solution correlates with Dimension 2 from the Caucasian MDS solution ($r(38) = .983, p < .01$), and Dimension 3 from the Asian MDS solution correlates directly with Dimension 3 from the Caucasian MDS solution ($r(38) = .892, p < .01$).

The element(s) represented by Dimension 3 was not as visually obvious and easy to identify as were race (Dimension 1) and gender (Dimension 2). Dimension 3 corresponds to the darkness of a face’s eyebrows and the darkness of a face’s irises. Both eyebrow darkness and iris darkness on each face were determined by two independent raters. The raters evaluated and coded these cues on a 3-point scale of light, medium, and dark. For Asian and Caucasian participants respectively, the correlations between Dimension 3 and both eyebrow darkness and iris darkness are: eyebrow darkness ($r(38) = .508, p < .01$; $r(38) = .514, p < .01$) and iris darkness ($r(38) = .474, p < .01$; $r(38) = .460, p < .01$). Additionally, two physical measurements from the faces correlate significantly with Dimension 3. Those two measurements are: 1) the width of the nose at the nostrils and 2) the distance between the base of the nose and top of the upper lip. The correlations between each of these two measurements and Dimension 3 for Asians and Caucasians, respectively, are: 1) nose width ($r(38) = .303, p < .05$; $r(38) = .364, p < .05$) and 2) length between nose and upper lip ($r(38) = .343, p < .05$; $r(38) = .307, p < .05$). Further, the ratio between these two measurements (nose width and the distance between the nose and mouth) correlates with Dimension 3 for both Asians ($r(38) = .428, p < .01$) and Caucasians ($r(38) = .361, p < .05$).
Dimension 4 from the Asian MDS solution correlates most strongly with Dimension 6 from the Caucasian MDS solution, $r(38) = .694, p < .01$. This suggests that the element Asians find fourth most important when judging facial similarity is the sixth most important element for Caucasians. Though we cannot say that Dimension 4/6 (Asian/Caucasian), as it will be referred to henceforth, captures any one element of the face, we can say that it refers to the length of facial features. For example, Dimension 4/6 correlates most strongly with the length of the chin, the ratio between forehead length to chin length, as well as the ratio between the length of the top half of the face to that of the bottom half, as divided at the base of the nose. The correlations between Dimension 4/6 and those measurements for Asians and Caucasians are: chin length ($r(38) = .590, p < .01; r(38) = .387, p < .05$), ratio between forehead length and chin length ($r(38) = .456, p < .01; r(38) = .356, p < .05$), and ratio of top to bottom half of the face ($r(38) = .598, p < .01; r(38) = .315, p < .05$).

We were able to identify the correlate to Dimension 6 from the Asian MDS solution as the presence or absence of a visible crease in the upper eyelid (distinguishing between “single eyelids” and “double eyelids”), $r(38) = .570, p < .01$. Asian faces naturally vary along this Dimension much more than do Caucasian faces (Caucasian faces nearly always have such an eyelid crease) just as Caucasian faces exhibit greater variability in the iris darkness (Asian faces nearly always have dark irises). The primary difference is that while iris color is captured by Dimension 3 for both Asians and Caucasians, indicating that both races deem this cue to be equally important when evaluating the similarity between all faces in our stimulus set, the presence or absence of
an eyelid crease is noticed only by Asians, according to the six dimensions that were extracted with the current multidimensional scaling techniques.

Dimension 5 from both the Asian and Caucasian MDS solutions as well as Caucasian Dimension 4 were not interpretable visually or mathematically using the physical measurements.

**Interracial Contact**

We also investigated whether one’s level of interracial contact influenced his similarity ratings of the faces and whether those individuals with greater interracial contact would produce a different MDS solution than those individuals with little interracial contact. Overall, the Asian participants reported a significantly greater amount of contact with Asians, and Caucasian participants reported a significantly greater amount of contact with Caucasians, $F(12,150) = 8.02, p < .001$ (See Question 5 from the Interracial Contact Questionnaire in Experiment 3, Figure 2). Analyzing the demographic data on participants’ contact with their three closest friends (See Question 6 from the Interracial Contact Questionnaire in Experiment 3, Figure 2), we found no significant results; although Asians report a greater number of Asian than Caucasian friends, they have not known their Asian friends for a longer period of time nor do they currently spend more time with their Asian friends than Caucasian friends. While Caucasian participants report more Caucasian than Asian friends and have known their Caucasian friends for a slightly longer period of time ($\chi^2(3) = 9.26, p < .05$), they do not spend more time with either their Caucasian or Asian friends.

Notably, for Asian participants, the zip code from the area where they lived the longest in life (See Question 4 from the Interracial Contact Questionnaire in Experiment
3, Figure 2) was the factor most associated with the amount of self-reported interracial contact with Caucasians; for Caucasian participants, there was no such factor that separated them into groups of high or low interracial contact. Demographic data was obtained on this zip code for each participant. If the Asian participant spent most time in this area where the proportion of Caucasians outweighed the proportion of Asians, this participant was coded as a high-contact (with Caucasians) Asian participant (N = 72). Otherwise, the participant was considered a low-contact (with Caucasians) Asian (N = 36). Interestingly, the proportion of Asians or Caucasians in the zip code where each participant has lived the longest is not related to the proportion of Asians or Caucasians reported as one of the participants’ three closest friends, χ²(3) = 2.89, p > .05.

Separate ALSCAL analyses were completed for high-contact and low-contact Asian participants. The six dimensions resulting from the high-contact Asian MDS solution correlate more strongly with those from the Caucasian MDS solution than do the six dimensions from the low-contact Asian MDS solution (see Experiment 3, Table 2). Recall that Dimension 3 from the Asian MDS solution correlates directly with Dimension 3 from the Caucasian MDS solution (r(38) = .892, p < .01) and that this dimension correlates with the darkness of the eyebrows and irises of the test faces. Dimension 3 from the high-contact Asian MDS solution correlates more strongly with Dimension 3 from the Caucasian MDS (r(38) = .923, p < .01) solution than does Dimension 3 from the low-contact Asian MDS solution (r(38) = .809, p < .01). Also recall that Dimension 4 from the Asian MDS solution correlates directly with Dimension 6 from the Caucasian MDS solution, r(38) = .694, p < .01 and that this dimension corresponds to the length of the chin, the ratio between forehead length to chin length, as well as the ratio between the
length of the top half of the face to that of the bottom half, as divided at the base of the nose. Notably, Dimension 4 from the high-contact Asian MDS solution correlates more strongly with Dimension 6 from the Caucasian MDS solution ($r(38) = .719$, $p < .01$) than does Dimension 4 from the low-contact Asian MDS solution ($r(38) = .582$, $p < .01$).

These findings suggest that Asians who had more contact with Caucasians perceived the stimulus set of faces more similarly to the Caucasian participants than did those Asians who had less contact with Caucasians.

**Discussion**

**Face Dimensions**

Using multidimensional scaling techniques, we are better able to understand the dimensions used by both Asians and Caucasians in judging the similarity and discriminating between faces. Firstly, the dimension of race is more important than that of gender for both Asians and Caucasians. Though this finding may surprise some, it is in line with findings across the eyewitness and face recognition literatures; there is a consistently noted own-race bias or cross-race effect (Kassin et al., 2001), while an own-gender bias is rarely discussed in the current literature. Though researchers have found that people are indeed better at recognizing faces of their own gender (e.g. Shaw & Skolnick, 1999; Wright & Sladden, 2003), Wright and Sladden revealed that hair accounts for half of this gender bias. Researchers have not similarly determined any one cue that generally accounts for a significant proportion of the own-race bias. If race is really a more important dimension than that of gender, one might believe that this dimension would correlate with a greater number of physical aspects of a face than would the dimension of gender. This is in fact true; for both Asians and Caucasians the
dimension of race correlates with more \((N = 11\) and \(N = 12\), respectively) physical measurements obtained on the faces in the experiment and ratios between these measurements than does the dimension of gender \((N = 9\) and \(N = 11\), respectively).

Interestingly, the fourth dimension from the Asian MDS solution correlates with the sixth dimension from the Caucasian MDS solution, indicating that participants of the two races see this dimension as having differential importance. Recall that this Dimension 4/6 correlates with the length of the chin, the ratio between forehead length to chin length, as well as the ratio between the length of the top half of the face to that of the bottom half, as divided at the base of the nose. The Asian participants place more emphasis on this dimension than do the Caucasian participants, though both use the dimension. This particular finding responds to an issue raised in the discussion of Experiments 1 and 2. We cautioned that cues should not be considered as either appropriate or inappropriate for a given race, but rather thought of on a continuum of more appropriate to less appropriate. Dimension 4/6 is direct evidence for this suggestion and may be considered more appropriate for Asian faces and less appropriate for Caucasian faces.

Somewhat consistent with our predictions, iris color/darkness emerged as an element of Dimension 3 on both Asian and Caucasian MDS solutions. Based on the results from the face description task in Experiment 1, we predicted that if this dimension emerged on the MDS solution for both races, it would emerge as a more important dimension for the Caucasian participants. This did not happen. One possible explanation for this inconsistency between our predictions and results lies in the fact that we were asking participants to compare not just two Asian faces (as iris color does not vary across
Asians, it is not a diagnostic cue for discrimination) or two Caucasian faces (as iris color does vary across Caucasians, it is a diagnostic cue for discrimination), but one face from each race as well. As such, all participants may have recognized that iris darkness was an important facial cue to use if: 1) they were trying to discriminate between two Caucasian faces or 2) they were trying to determine the race of a test face (recall that we removed any external race markers with the morphed hairline). We have already discussed the fact that Caucasian faces vary along the dimension of iris color/darkness far more than do Asian faces. Identification of the iris darkness may have been an important cue for participants determining whether a face was Asian or Caucasian and subsequently determining the similarity rating for a pair of faces.

Regarding the feature of eyelid crease, we predicted that if this feature emerged as a dimension on an MDS solution, it would do so only for the Asian participants. Results are entirely in line with this hypothesis. The presence/absence of an eyelid crease correlates with Dimension 6 for the Asian participants, but does not correlate with any of the six dimensions from the Caucasian participants’ MDS solution. This is an appropriate cue for Asians to use when judging the similarity between two faces because of the natural variability in this feature on Asian faces. These findings are consistent with the findings from Experiment 1 that only Asian participants referenced single eyelids (absence of a crease) or double eyelids (presence of a visible crease) when describing faces.

Interracial Contact

Researchers have used many different measures to assess the quantity and quality of interracial contact experienced by their participants, as we noted in our discussion
about the contact theory of the ORB. The two most predominant methods for collecting this evidence has been through gathering population statistics (e.g. Aronstam & Tyson, 1980; Platz & Hosch, 1988) as well as obtaining self-reports on quantity of cross-race associations and the frequency of contact (e.g. Malpass & Kravitz, 1969; Brigham & Barkowitz, 1978). Investigators have only collected population statistics on the area where they gather their data and where participants currently live, but have not yet used population demographics about the area where participants lived the longest in life (and potentially had more or less interracial contact than where they live at the time they participate in the research). The present experiment enhances the literature in that we used a combination of novel demographic data and self-report measures to assess our participants’ interracial contact.

We found a relationship between participant race and the self-reported amount of contact with same- and other-race individuals such that both Asians and Caucasians reported spending more time with other people of their own race. Asians who lived the longest in areas where the proportion of Asians outweighed the proportion of Caucasians (low-contact Asians) reported significantly less interracial contact than did those Asians who lived the longest in areas where the proportion of Caucasians was greater (high-contact Asians). Unfortunately, we found no measure that separated the Caucasian participants into high- and low-contact groups. Although high-contact Asians did not differ significantly from low-contact Asians in respect to their MDS solutions, the dimensions from the low-contact Asian MDS solution correlated more strongly (than did the dimensions from the high-contact Asian MDS solution) with the dimensions from the Caucasian MDS solution (see Experiment 3, Table 2).
Of particular interest is that Dimension 3 for high-contact Asians correlated more strongly with Dimension 3 for the Caucasians than did Dimension 3 for the low-contact Asians. Recall that iris color/darkness loads on to Dimension 3 for both races. We predicted that iris color would be a more important dimension for Caucasians than for Asians, and though we did not find this difference, results suggest that iris color may be a slightly more important dimension for high-contact Asians than for low-contact Asians. Similarly, it is of interest that Dimension 4 for high-contact Asians correlated more strongly with Dimension 6 for the Caucasians than did Dimension 4 for the low-contact Asians. Overall, the similarity judgments for the pairs of test faces were perceptually more similar for high-contact Asians and Caucasian participants than for low-contact Asians and Caucasians. These findings suggest that increased interracial contact can lead to learning the appropriate cues for perceptually discriminating faces.

A Combination of Theories

Results from the present experiment support a combination of theories explaining the own-race bias, namely the contact hypothesis and the inappropriate cue utilization hypothesis. Researchers have been struggling to embrace a unified theory for the ORB (Meissner & Brigham, 2001), in part because many studies investigate only one theory at a time. Through multidimensional scaling and the collection of interracial contact information, we were able to assess two prominent theories in conjunction with one another. Several clear implications have emerged from this study. Firstly, different races do in fact use different cues when evaluating faces. Experiment 1 suggested that different races use different cues when describing faces, but we were unable to determine if those cues that people use to describe faces are those cues that people actually use.
when discriminating faces. Researchers have previously demonstrated this using different techniques such as face description tasks (e.g. Doty, 1998), but this is the first study to use MDS and to compare MDS solutions for different race participants on the same stimulus set of mixed-race faces. The claim that different races place differential importance on different cues is evident from the fact that not all six dimensions from the Asian MDS solution correlate directly with all six dimensions from the Caucasian MDS solution. Demonstrating this effect in an implicit perceptual task (e.g. inferring the appropriate cues from MDS on similarity ratings) adds to the current literature that has only clearly demonstrated this in explicit perceptual tasks (e.g. participants directly reporting the most appropriate/useful cues).

Secondly, even when races use similar cues, they do not necessarily prioritize those cues in the same order. The fact that Dimension 4 from the Asian MDS solution correlates with Dimension 6 from the Caucasian MDS solution is evidence for this claim. The cues that Asians find fourth most important in rating the similarity between faces is only sixth most important for Caucasians. The inappropriate cue utilization hypothesis predicts that we use those cues most appropriate for discriminating same-race faces when discriminating both same- and other-race faces. Results from the present study suggest that the dimensions revealed for Asian and Caucasian MDS solutions are not only those dimensions that Asians and Caucasians use when discriminating all faces, but are also those dimensions that are most appropriate for discriminating same-race faces. For example, Dimension 4/6 may be more appropriate for the perceptual discrimination of Asian faces (as it is a higher dimension on the Asian MDS solution) and less appropriate
(though not inappropriate) for Caucasian faces (as it is a lower dimension on the Caucasian MDS solution).

Lastly, the quantity/quality of interracial contact, as discussed above, does appear to affect at some level the cues we use to process faces. The combination of these findings provides evidence that the own-race bias cannot be explained purely by differential contact or by the inappropriate cue utilization hypothesis, but by an interaction between the theories.
CHAPTER 5

Experiment 4

Introduction

Experiments 1 and 3 provided evidence that people from different races do in fact use different cues to describe and distinguish between faces. Across both experiments, we found that the presence/absence of an eyelid crease is a more important cue to distinguish Asians than to distinguish Caucasians and indeed Asians use this cue more often than do Caucasians, who do not appear to use this cue at all. This disparity may be due to the fact that the presence/absence of an eyelid crease is a more variable cue for Asians than for Caucasians (who nearly always have a crease in the upper eyelid) and thus more diagnostic for recognizing Asian faces. Results from Experiment 2 support this idea and demonstrate that people in general understand the fact that Asians vary more along the dimension of eyelid crease than do Caucasians. We are then left with the question: if Caucasians learn that eyelid crease is an important dimension along which Asian faces may be discriminated, can we lessen the extent to which they exhibit the own-race bias?

Results from the multidimensional scaling analysis in Experiment 3 suggest that interracial contact can influence the cues people use to discriminate faces. The literature surrounding the contact hypothesis of the ORB implies that merely increasing Caucasians’ contact with Asians will not necessarily increase their use of the eyelid crease cue and subsequent ability to discriminate Asian faces (e.g. Furl, Phillips, & O’Toole, 2002; Brigham & Malpass, 2001; Levin, 2000). Levin (2000) states that “mere contact is not sufficient to eliminate the CR [cross-race] recognition deficit. Instead, the
perceiver must process CR faces with the goal of individuation (p. 560).” It is clear, however, that the ORB can be reduced or even eliminated given sufficient experience, motivation and practice. For example, Sangrigoli et al. (2005) examined the recognition ability of Korean adults who had been adopted at an early age (between ages 3 and 9) by Caucasian families. Comparable to a group of Caucasian participants who had also been raised in Caucasian homes, these particular Korean adults were better able to identify Caucasian than Asian faces. These findings demonstrate that the ORB can be not only mitigated, but even reversed given ample motivation and experience. Hugenberg, Miller, and Claypool (2006) used verbal (not visual) training and demonstrated that when Caucasian participants are made aware of the ORB phenomenon or are asked to individuate faces (as suggested by Levin (2000)), participants were less likely to exhibit the ORB when viewing Black and White faces in a traditional face recognition paradigm. The researchers told participants that they should look for features that best discriminated between individuals of the other-race, though they did not give participants information as to the specific features that actually do discriminate best between other-race faces. Perhaps with increased experience and motivation, these participants could automatically make these finer discriminations between not only same-race faces but other-race faces as well.

Is it possible to decrease the ORB if we visually train people on those cues most important for discrimination? Research from training tasks using both human and non-human visual stimuli suggests that this is entirely possible. Gauthier and Tarr (1997) trained expertise in “Greebles,” a fictional domain in which no experts (or individuals with more experience) naturally exist. They trained experts to categorize Greebles (into
gender groups, family groups, or individualization) based on three appropriate features. Consistent with reported performance of experts in other domains of the literature, participants who developed Greeble expertise were reliably more accurate at discriminating Greebles and were more sensitive to inversion (e.g. Diamond & Carey, 1986) of a studied Greeble than were novices. Gauthier and Tarr’s (1997) findings raise the question of whether we can train people on the appropriate cues for discriminating not only fictional groups, but for discriminating cross-race faces for which people do not have experience and expertise.

Elliott, Willis, and Goldstein (1973) found that practice with perceptually discriminating Asians improves Caucasians’ subsequent discrimination ability for Asian faces. Caucasian subjects were assigned to one of three conditions: 1) practice discriminating Asian faces, 2) practice discriminating Caucasian faces, or 3) no practice with either race. Those participants who had practice with Asians showed better subsequent recognition of Asians than participants in the other two groups. This study speaks directly to the contact theory of the ORB in that increased contact with Asians improved recognition ability. However, these results suggest but do not allow us to definitively understand whether people were learning the appropriate cues for discriminating faces or whether these findings are a result of practice effects. Hills and Lewis (2006) trained White participants to distinguish Black faces using the cues that are most important for recognition of Blacks. They based their training on findings from Shepherd and Deregowski’s (1981) study examining descriptions of White and Black faces. Shepherd and Deregowski found that the lower facial features (e.g. mouth, nose, chin, and cheeks) account for 75% of the variance in Black faces. Hills and Lewis then
used these results to create their training task which trained Caucasians to discriminate
composite faces that differed only on the lower facial features (mouth, nose, chin, and
cheeks). Interestingly, participants were not trained to discriminate Black faces that
naturally vary along those dimensions, but were trained to discriminate same-race White
faces that were altered on the basis of their lower facial features. Whites who learned to
discriminate faces based on the cues most diagnostic for Black faces demonstrated a
reduced ORB (superior recognition of Black faces) compared to those Whites who
learned to discriminate faces based on cues not diagnostic for Black faces.

In the present study, we were interested in whether training Caucasian participants
to discriminate faces on the basis of eyelid crease would mitigate the own-race bias with
respect to Asian faces. Rather than training participants using composites of same-race
faces, we use actual photos of same-race and other-race faces in the training. The entire
experimental task is broken into three segments: 1) We assess participants’ baseline ORB
with a traditional face recognition paradigm, 2) We then train participants to attend to a
race-appropriate cue for Asian faces or a race-irrelevant cue, and 3) Participants’ post-
training ORB is then assessed using a second standard face recognition paradigm.
Participants are randomly assigned to one of three experimental or control training groups
that train participants to attend to: 1) presence/absence of eyelid creases, 2)
lightness/darkness of eyebrows, or 3) face race. The eyebrow training and race training
conditions are both intended as control groups. In particular, the darkness of eyebrows is
appropriate for both Asian and Caucasians; multidimensional scaling analyses in
Experiment 3 indicated that both Asians and Caucasians attend to eyebrow darkness
equally. Also, training participants to attend to an irrelevant area of the eye region
provides a comparison for the eyelid crease training condition. The race training condition teaches both Asians and Caucasians to attend to the race of own-race and cross-race faces. MDS analyses in Experiment 3 also revealed that both races attend to race equally, making the race training a valid control training.

We predict that both the eyebrow training and race training should have no effect on participants’ ORB. The eyelid crease training, however, should improve Caucasians’ recognition ability for Asian faces. This training should not affect Caucasians’ recognition of Caucasian faces or Asians’ recognition of either race faces. The eyelid crease training is the only training condition that contains one cue that is more appropriate for discriminating between faces of one race than the other race, and for this reason it is the only training that should affect participants’ post-training ORB.

Method

Participants

All participants (N = 550) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 31% male, 26% Caucasian, and 74% Asian.

Stimuli

The stimuli consisted of 80 pictures (225 x 311 pixels, gray scale): 40 Caucasian men and 40 Asian men. The original photographs were provided by Dr. Jim Tanaka and are the same photos used in Experiment 2. Each of the 40 Asian faces had been placed inside an identical hairline with the same background; the same was true for the 40 Caucasian faces. All faces had neutral expressions and contained no facial hair or jewelry (see Experiment 4, Figure 1 for examples of the test stimuli). Additionally,
participant responses were collected on the keyboard using the numeric keys. Stimuli presentation as well as participant response data were controlled and collected by an E-Prime script.

Procedure

The procedure consisted of three primary segments. Participants first completed a face recognition task followed by a training task and then a subsequent face recognition task. Participants were randomly assigned to one of three groups, differing on the content of the training task.

In the first recognition task, participants viewed a set of 20 faces: 10 Caucasian men and 10 Asian men (half with single eyelids and half with double eyelids). These faces were presented for three seconds each, in random order for each participant. Participants then completed a short distracter task where they saw old yearbook photos of famous celebrities and were asked to silently name each face (no response was required). They then viewed a set of 40 faces: 10 old Caucasian men, 10 new Caucasian men, 10 old Asian men, 10 new Asian men (half with single eyelids and half with double eyelids). Each face was presented for three seconds and then remained on the screen while participants indicated whether the face was old (they had seen it in the first set of faces) or new (they had NOT seen it in the first set of faces).

The training task had the same structure for all participants, but the focus of the task differed between the three groups of participants. We will refer to these tasks as “eyelid training,” “race training,” and “eyebrow training,” referencing the focus of the trainings. The details of these trainings will be discussed with specific reference to the eyelid training group. Participants were told that they were learning the difference
between single eyelids and double eyelids (or between light and dark eyebrows in the eyebrow training and between Asians and Caucasians in the race training). They were not given any information about the purpose of the training, however. Participants were first shown two Caucasian faces and two Asian faces (all four from the initial face recognition task). One face from each race contained single eyelids and the other face had double eyelids. Participants were also given descriptions of single and double eyelids along with the photos. These descriptions referenced the presence or absence of a visible crease in the upper eyelid as the critical distinction between double and single eyelids.

The actual training task was broken into two segments. The first segment presented participants with two faces and asked them to indicate whether the faces were the same or different, based solely on the type of eyelids on the two faces. Faces were considered the same if both faces had single eyelids or if both faces had double eyelids. All the faces used in the training task were those 40 faces (20 Asian, 20 Caucasian) used in the initial face recognition task. Participants answered this question for 40 pairs of faces (10 pairs of same-race Asian faces, 10 pairs of same-race Caucasian faces, and 20 pairs of cross-race faces) presented simultaneously side by side in random position (left/right) and order. Because we only used 40 faces in the training, participants saw the faces several times throughout the training. Participants received feedback after each response indicating whether their response was correct/incorrect as well as the reason.

The second segment of the training task presented participants with a single face and asked them to indicate whether the face had single eyelids or double eyelids. Participants answered this question for each of the 40 faces presented in random order. Again, they
received feedback after each response. Participants then repeated both segments of the training task.

The eyebrow training condition trained participants to discriminate faces on the basis of lightness/darkness of the eyebrows, and the race training condition trained participants to discriminate faces on the basis of race (Asian vs. Caucasian).

After completing the training task, all participants completed a second face recognition task with 40 faces not previously seen in the first face recognition task or in the training. The structure of the recognition task was identical to that of the first recognition task.

In summary, the procedure started with a face recognition task. Participants then underwent training where they learned to discriminate faces on the basis of: presence/absence of eyelid crease, eyebrow lightness/darkness, or race. The training consisted of 2 components: 1) participants determining whether two faces were the same or different on the basis of the trained dimension in the assigned condition and 2) participants rating each face on the basis of the trained dimension. After the training, participants completed a second face recognition task in order to assess whether the training task affected recognition of Asian and Caucasian faces.

After finishing this computer-based task, participants completed a brief survey (identical to that used in Experiments 1 and 2) of personal demographic information including zip codes of where participants were born and of where they have lived the longest in life.

Data Analysis
We measured participants’ ORB using the signal detection measure of $d'$, as calculated by: $d' = \text{proportion of correct identifications (hits)} - \text{proportion of incorrect false identifications (false alarms)}$. This $d'$ is a measure of discriminability, or discrimination performance. For each participant, we calculated $d'$ for same-race faces and for other-race faces both pre-training and post-training. From these calculations, we were able to determine the change in $d'$ for same-race and other-race faces.

We also calculated each participant’s response criterion/response bias, or $c$, as calculated by: $c = z(\text{hits}) + z(\text{false alarms})$. $c$ is bounded at -1 and 1 with a neutral point at 0 (where signal and noise distributions cross), and is measured in standard deviation units. The criterion indicates how likely a participant was to respond “yes” or “no.” If the criterion is high, then the participant was more likely to say “no, I haven’t seen that face before.” On the other hand, if the criterion is low, then the participant was more likely to respond, “yes, I have seen that face before.” As with $d'$, we calculated each participant’s $c$ for same-race and other-race faces, both pre-training and post-training. Also like $d'$, we were able to calculate the change in $c$ for same- and other-race faces.

Results

Pre-training recognition

We evaluated all participants’ pre-training own-race bias with a 2 x 2 ANOVA with participant race (Asian, Caucasian) as a between-subjects independent variable, test face race (Asian, Caucasian) as a within-subjects independent variable, and $d'$ as the dependent variable. There was a main effect of test face race such that participants exhibited a higher $d'$ for Asian faces ($M = 1.30$) than for Caucasian faces ($M = 1.17$), $F(1,549) = 4.58, p < .05$. There was also a trend (but no significant result) toward an
interaction between participant race and face race \((F(1,549) = 2.47, p > .05)\) such that Asian faces were better discriminated by Asian \((M = 1.37)\) than by Caucasian participants \((M = 1.24)\) and Caucasian faces were better discriminated by Caucasian \((M = 1.20)\) than by Asian participants \((M = 1.14)\). This combination suggests (without statistical significance) that participants in our sample exhibited the own-race bias prior to training. Concrete evidence for this pre-training ORB comes from examining participants’ criterion bias, or \(c\). We conducted a 2 x 2 ANOVA using participant race and test face race as independent variables and \(c\) as the dependent variable and found a significant main effect of participant race as well as a significant interaction between participant race and test face race. In particular, Asians showed a greater bias toward saying “no, I have not seen that face before” \((M = .09)\) and Caucasians were less likely to respond “no, I have not seen that face before” \((M = .01)\), \(F(1,549) = 4.54, p < .05\). Additionally, Asian participants were more likely to respond that they had seen a Caucasian face \((M = -.01)\) before than an Asian face \((M = .19)\), and Caucasian participants were more likely to say that an Asian face \((M = -.04)\) was “old” than for a Caucasian face \((M = .06)\), \(F(1,549) = 14.79, p < .001\). This interaction is evidence of the ORB such that Asians were more likely to false alarm on Caucasian than on Asian faces by setting a more liberal response criterion for Caucasian faces, and Caucasians were more likely to false alarm on Asian faces by setting a more liberal response criterion for Asian faces (see Experiment 4, Graph 1).

Post-training recognition

A 3 x 2 x 2 ANOVA was conducted using condition (eyelid training, eyebrow training, and race training) and test face race (Asian, Caucasian) as within-subjects
independent variables, participant race (Asian, Caucasian) as a between-subjects independent variable, and the change in $d'$ (from pre-training to post-training) as the dependent variable. We found a significant main effect of condition on $d'$ change ($F(2,548) = 4.21, p = .01$), such that participants in the race training condition ($M = .20$) demonstrated a significant increase in $d'$ post-training over those participants in either the eyelid training ($M = -.09$) or eyebrow training conditions ($M = -.02$). Participants (both Asian and Caucasian) in the race training condition showed increased discriminability for all faces (Asian and Caucasian) post-training compared to pre-training. There were no other significant main effects or significant interactions. We also conducted a Repeated-Measures ANOVA with condition and test face race as within-subjects independent variables, participant race as a within-subjects independent variable, and $d'$ as the repeated measure ($d'$ pre-training and $d'$ post-training) dependent variable. Analyzing the data in this manner, we replicated the results from the original ANOVA, finding only a significant main effect of condition on $d'$.

A second 3 x 2 x 2 ANOVA was conducted using condition (eyelid training, eyebrow training, and race training) and test face race (Asian, Caucasian) as within-subjects independent variables, participant race (Asian, Caucasian) as a between-subjects independent variable, and the change in $c$ (from pre-training to post-training) as the dependent variable. We found a significant main effect of test face race on $c$ change, $F(2,548) = 5.53, p < .05$. Across conditions and participant race, participants demonstrated a positive shift in $c$ post-training for Asian faces ($M = .04$) and a negative shift in $c$ post-training for Caucasian faces ($M = -.06$). Post-training, participants set a more conservative criterion for Asian faces compared to pre-training and set a more
liberal criterion for Caucasian faces compared to pre-training. There were no other significant main effects or interactions. In a different analysis of the decision criterion, we conducted a Repeated-Measures ANOVA with condition and test face race as within-subjects independent variables, participant race as a between-subjects independent variable, and \( c \) as the repeated measure (\( c \) pre-training and \( c \) post-training) dependent variable. Analyzing the data with this method, we replicated the results from the original ANOVA, finding only a significant main effect of test face race on \( c \).

Because there was no main effect of training condition on decision criterion (\( c \)), we ran a 2 x 2 ANOVA collapsing across training condition with participant race as a between-subjects independent variable, test face race as a within-subjects independent variable, and the average (between pre-training and post-training) \( c \) as the dependent variable. In this analysis, we found significant main effects of participant race and test face race as well as a significant interaction between participant race and test face race on \( c \). Asian participants set a higher criterion (\( M = .09 \)) than Caucasian participants (\( M = - .002 \)), \( F(1,549) = 10.21, p = .001 \), indicating that Asian participants were less likely to respond that they had previously seen a test face. Further, all participants set a higher criterion for Asian faces (\( M = .10 \)) than for Caucasian faces (\( M = -.01 \)), \( F(1,549) = 7.53, p < .01 \), an indication that participants were more likely to respond that they had previously seen a Caucasian face than an Asian face. Evidence of the ORB lies in the interaction between participant race and test face race; across pre- and post-training recognition tasks, Asian participants set higher decision criterion for Caucasian faces than for Asian faces, and Caucasian participants set higher criterion for Asian than Caucasian faces, \( F(1,549) = 20.13, p < .001 \) (see Experiment 4, Graph 2). By setting a higher, more
liberal criterion for other-race faces, participants are more likely to false alarm on cross-race than on same-race faces.

**Interracial Contact**

We were curious as to whether one’s level of interracial contact influenced recognition of both Asian and Caucasian faces. In a pilot study as well as in Experiment 3, we found that the strongest marker for Asian participants’ levels of interracial contact with Caucasians was the zip code from the area where they reported living the longest in life. Therefore, this was the demographic data used to divide Asian participants into high- and low-contact groups (as previously done in Experiments 1, 2, and 3). Asians spending most time in an area where the proportion of Caucasians outnumbered the proportion of Asians were considered high-contact (63%). Otherwise, the participant was considered low-contact (37%). Because only two Caucasians in this sample reported living the longest in an area where the proportion of Asians was greater than the proportion of Caucasians, we once again constrain our interracial contact analyses solely to the Asian participants.

To see if level of interracial contact affected discrimination ability, we first conducted a 2 x 2 ANOVA with level of interracial contact as a between-subjects independent variable, test face race as a within-subjects independent variable, and pre-training $d'$ as the dependent variable. We did not find a significant interaction between level of interracial contact and test face race on $d'$, as would be predicted by the contact hypothesis, $F(1,406) = .01, p > .05$. Level of interracial contact did not affect Asians’ abilities to differentially discriminate Asian or Caucasian faces. We found a significant main effect of test face race on $d'$ such that all Asians (regardless of level of contact)
exhibited superior discrimination of Asian faces ($M = 1.33$) than of Caucasian faces ($M = 1.04$), $F(1,406) = 19.05$, $p < .001$.

To see if level of interracial contact affected response criterion/bias, we conducted a 2 x 2 ANOVA with level of interracial contact as a between-subjects independent variable, test face race as a within-subjects independent variable, and pre-training $c$ as the dependent variable. We did not find a significant interaction between level of interracial contact and test face race on $c$, as would be predicted by the contact hypothesis, $F(1,406) = .35$, $p > .05$. Interracial contact did not differentially affect Asians’ response bias for Asian and Caucasian faces. We did find a significant main effect of test face race on Asian participants’ criterion such that all Asians set a higher, more conservative criterion for same-race Asian faces ($M = .18$) than for Caucasian faces ($M = -.02$), $F(1,406) = 23.72$, $p < .001$. Interestingly, we also found a main effect of level of interracial contact on criterion such that high-contact Asians ($M = .04$) set lower, more liberal criteria for faces (both Asian and Caucasian) than did the low-contact Asians ($M = .13$), $F(1,406) = 4.54$, $p < .05$. High-contact Asians were more likely to report that they had previously seen a test face than were low-contact Asians.

Discussion

Recall that the training task was intended primarily to alert Caucasian participants to the Asian race-appropriate cue of eyelid crease, though both Asians and Caucasians underwent the training. We measured participants’ recognition abilities in terms of $d'$ as well as $c$. With respect to $d'$, the different training conditions did not differentially affect Caucasian participants’ own-race bias, as was hypothesized. Recall that although the pre-training interaction between participant race and test face race on $d'$ was not significant,
the trend was such that both Asians and Caucasians showed better discriminability for same-race faces than for cross-race faces. We predicted that Caucasians would show increased discrimination ability for Asian faces after the eyelid training, and that the eyebrow training as well as race training would serve as control conditions and fail to affect Caucasians’ subsequent recognition of Asian faces. In fact, results from the eyelid training mirrored the results from the eyebrow training (intended as a control condition). Participants’ discriminability for Asian and Caucasian faces did not change after the eyelid or eyebrow training. Interestingly, participants in the race training demonstrated increased (d’) for both Asian and Caucasian faces after training. There are two simple explanations for this finding: 1) the race training condition alerted participants to the race of the faces and perhaps increased their sensitivity and awareness to within-race differences in the faces, and 2) the increased sensitivity may be a straightforward demonstration of practice effects. Practice effects support the contact hypothesis behind the ORB in that simply via exposure individuals can gain awareness of the differences between races. Although participants in all three conditions had equal amounts of contact with the training and test faces (seemingly nullifying any explanation of differential practice effects), the eyelid training and eyebrow training focused participants on a specific aspect of the face, perhaps precluding attention and exposure to the entire face (thus providing participants in the race training with differentially more exposure). This could explain the unexpected result of increased discriminability for participants only in the race training condition.

Examining recognition ability in terms of c, participants exhibited the own-race bias before training, evident by the statistically significant result that Asians set a more
liberal decision criterion for Caucasian than for Asian faces and Caucasians conversely set a more liberal decision criterion for Asian than for Caucasian faces. Participants were less likely to indicate that a same-race face was “old” or that they had seen this same-race face before; therefore participants were less likely to false alarm on same-race faces and more likely to false alarm on cross-race faces for which they had placed a more liberal decision criterion. After undergoing any one of the three training tasks, participants in the second recognition task exhibited an increase in decision criterion bias, $c$, for Asian faces and a decrease in $c$ for Caucasian faces. Participants became more conservative when responding to Asian test faces and more liberal when responding to Caucasian test faces. This trend appears to be driven by a larger increase in $c$ for Asian faces in the eyelid training and a larger decrease in $c$ for Caucasian faces in the race training condition; perhaps participants in the eyelid training condition were sensitized to the fact that Asian faces vary more than Caucasians on the dimension of eyelid crease and thus set more conservative criterion for Asian faces in the second recognition task.

It is likely that sensitivity and perceptual expertise cannot be gained in an hour’s worth of training. Future research should include a more long-term training to develop perceptual expertise for the appropriate cues needed to make more accurate cross-race discriminations. Further, it may be advantageous to train participants on the appropriate cues for cross-race face discrimination by varying and training those cues on own-race faces (rather than training on cross-race faces) as done by Hills and Lewis (2006). Consistent with the inappropriate cue utilization hypothesis that people apply those cues that they use for same-race discrimination to other-race discrimination, it seems that if participants learn to discriminate own-race faces on the basis of a given cue that is more
appropriate for other-race discrimination, they may then appropriately apply this cue to
other-race faces. Though we did not see the expected training effects on memory for
faces, this study leaves open the possibility that training on a specific feature may have
more visible effects on face perception. Recall that we discussed the concept of long-
term memory encoding items semantically in terms of the general idea (Craik & Tulving,
1975) and that facial features may be more important for perception than for memory.
Future research should include measuring face perception (e.g. using MDS as done in
Experiment 3) before and after training on a specific cue such as eyelid crease. It may
also be valuable to train participants to discriminate faces using more configural rather
than featural information (e.g. Hills and Lewis, 2006); this variation on the training task
may more effectively improve recognition memory for cross-race faces than did our
training task using a specific, featural cue.

There are several important enhancements to the own-race bias literature that this
study provides. For one, researchers typically measure the ORB solely with the signal
detection measure of $d'$, using correct identification (hit) rates, or using incorrect
identification (false alarm) rates (e.g. Hills & Lewis, 2006; Bernstein, Young, &
Hugenberg, 2007; Hugenberg, Miller & Claypool, 2006). Researchers currently do not
discuss the decision criterion $c$ with reference to the ORB. Looking at participants’
decision criterion provides a different angle for analyzing participant responses. In
particular, without information on this criterion, it must be assumed that the criterion is
neutrally set at 0 and that participants are perfectly calibrated to respond “yes” and “no”
with equal frequency to faces of both races. Evidence from the present study
demonstrates that this should not be the default assumption and that participants are more
conservative for same-race faces and more liberal on where they set their decision criterion for other-race faces.

Although we did not see the predicted effects of the training in Experiment 4, we were able to make both methodological as well as theoretical advancements in the analysis of the ORB. We provided a springboard from which researchers are now better equipped to investigate appropriate cue utilization as a means to reduce the own-race bias.
CHAPTER 6

General Discussion

In this paper we first theoretically discussed the mechanisms driving the own-race bias. To date, the literature has represented several competing, if not mutually exclusive, and complementary theories, including an inherent difficulty hypothesis, a prejudicial attitudes hypothesis, a differential processing hypothesis, an inappropriate cue utilization hypothesis, and a contact hypothesis, none of which have alone provided a thorough understanding of the phenomenon (e.g. Meissner & Brigham, 2001; Sporer, 2001; Ayuk, 1990). The four experiments we presented provide evidence for the inappropriate cue utilization hypothesis, in conjunction with the contact theory, as a foundational principle underlying the own-race bias. Additionally, these experiments provide further advancements regarding the exact nature of the theories along with methodological developments in the investigation of these theories.

Though we are not the first to suggest the importance of contact together with the importance of learning to discriminate using the appropriate cues, we advance the current literature in two important ways: 1) we use methodologies that are novel for examining the inappropriate cue utilization hypothesis as well as the contact theory, and 2) we update the inappropriate cue utilization hypothesis beyond its current general form. We discuss our results first in relation to the inappropriate cue utilization hypothesis and then in terms of the contact theory.

Inappropriate cue utilization hypothesis

Recall that in its most basic form, the inappropriate cue utilization hypothesis suggests that people use cues important for same-race face recognition and then apply
those cues inappropriately to discriminate other-race faces (Shepherd, 1981).

Experiments 1 and 3 provide evidence for the existence of cues that are indeed more
diagnostic for discriminating between members of one race than another. These
diagnostic cues have high within-race variability, important for discriminating between
faces of that race. Experiment 1 asked participants to provide descriptions of same- and
other-race faces with the expectation that a participant’s friend could pick the face out of
a crowd on the basis of the participant’s description alone. Results indicate that Asians
describe the presence or absence of an eyelid crease on a face more than do Caucasians,
while Caucasians describe the iris color on a face more than do Asians. The ideal
observer model (Geisler, 1989) predicts that people use those cues with the most
variability (and thus increased diagnosticity) to discriminate between faces; it follows
that iris color is a more variable (and thus more appropriate) cue for Caucasian faces and
that the presence/absence of an eyelid crease is a more variable (and thus more
appropriate) cue for Asian faces.

Additionally, Asians and Caucasians in Experiment 1 describe these race-
appropriate cues differentially more for faces of their own race than for faces of another
race. In its current form, the inappropriate cue utilization theory implies that people use
cues appropriate for same-race discrimination to distinguish same-race faces with the
same frequency as which they use those same cues to distinguish other-race faces. Our
results suggest, however, that at some level people may realize that these cues are more
appropriate for same-race face discrimination than for other-race face discrimination
because although they apply them to other-race faces, they do so at a significantly lower
rate than which they apply them to same-race faces. Nonetheless, it still leads to the
ORB. This is an important distinction because it clarifies the inappropriate cue utilization hypothesis (e.g. Shepherd, 1981) where it previously lacked specificity.

Further evidence for various cues being more appropriate for some races than others (though not necessarily dichotomously appropriate versus inappropriate) lies in the results from Experiment 3 which used multidimensional scaling to elucidate the facial features that Asians and Caucasians attend to most when comparing faces. Recall that Experiment 3 asked participants to rate the similarity of pairs of simultaneously presented Asian and Caucasian faces. We then performed MDS analyses using the average similarity rating for each pair of faces and compared the resulting MDS solutions for Asian and Caucasian participants. Out of six dimensions, MDS revealed that lengthening characteristics of the face (i.e. the ratio of the forehead to the chin and the ratio of the top half of the face to the bottom half) load on to the fourth most important dimension for Asians but only the sixth most important dimension for Caucasians. Furthermore, eyelid crease is the sixth most important dimension for Asians, but not attended to by Caucasians (or at least it is not evident in the six dimensions revealed from the Caucasian MDS solution). Interestingly, Asians and Caucasians attended to iris color equally in the MDS task; perhaps when comparing Asian and Caucasian faces, Asians may attend to this feature more than they do when only viewing Asians faces because they realize that it is an important cue for either: 1) discriminating an Asian face from a Caucasian face or 2) discriminating two Caucasian faces from one another. It is possible that a similar trend occurs for Caucasians when viewing both race faces. Caucasians may attend to eyelid creases more when viewing a set of Asian and Caucasian faces than when viewing only Caucasian faces, but it is a sufficiently insignificant dimension for Caucasians that it is
not captured by the six dimensions that our MDS techniques revealed. It is also
important to note that both Experiments 1 and 3 are the first face description and
multidimensional scaling tasks to use Asian and Caucasian participants evaluating Asian
and Caucasian test faces. Face description research has not before included Asian
participants or test faces, and MDS studies have not before separately evaluated the MDS
solutions for different race participants viewing the same set of mixed-race stimuli.

Experiments 1 and 3 suggest that there are significant perceptual differences
between own- and other-race faces, and we wanted to explore whether these differences
also exist in memory. Experiment 2 evaluated the strength of differentially appropriate
cues (iris color as more important for Caucasians and eyelid crease as more important for
Asians) in two recognition memory tasks rather than in more perceptual tasks such as
Experiments 1 and 3. This is a significant distinction because the literature has yet to
clarify whether the own-race bias is perceptually based or memory-based (e.g. Meissner
& Brigham, 2001). In Experiments 2a and 2b, we used traditional recognition memory
paradigms to evaluate whether people are sensitive to variations in iris color for Asian
faces (outside the expected range of variability) and Caucasian faces (within the expected
range of variability) as well as variations in the presence/absence of eyelid creases for
Asian faces (within the expected range of variability) and Caucasian faces (outside the
expected range of variability). In Experiment 2b, we altered cues only within their
expected ranges of variability (i.e. swapping iris color only on Caucasian faces and
swapping the presence/absence of eyelid creases only on Asian faces).

Though participants in Experiments 2a and 2b did demonstrate the own-race bias
with respect to unaltered faces, this bias disappeared when we altered iris color and the
presence/absence of eyelid creases both within and outside their natural ranges of variability. In particular, when we altered iris color outside its expected range for Asian faces, all participants were highly sensitive to this change. On the other hand, when we altered iris color within the range of variability for Caucasian faces and altered eyelid creases within the expected range for Asian faces, both races of participants were equally insensitive to these changes. While results from Experiments 1 and 3 imply significant perceptual difference in own- and other-race faces with respect to iris color and presence/absence of an eyelid crease, Experiment 2 suggests no significant difference in the memory of own- and other-race faces with respect to these race-appropriate cues.

The findings from Experiment 2 suggest that people have a basic understanding of the natural range of variability for same- and other-race face cues, but that a short memory task is insufficient for individuals to attend to these discriminating features, perhaps even for faces of their own-race. These results suggest that changes in race-appropriate facial features do not affect memory for same- or other-race faces. This is consistent with findings from the eyewitness literature that eyewitness identification can be extremely unreliable in general (e.g. Wells et al., 1998; Innocence Project, 2001), and even more fallible when the witness and suspect are of different races (Meissner & Brigham, 2001). Therefore, it is important that we understand the mechanisms driving the own-race bias such that we can better evaluate the reliability of eyewitnesses in actual cases.

If we uncover cues that are important for discriminating members of a given race, is it possible that we can train people to use those appropriate cues and hence mitigate or even eliminate the ORB? We asked this very question in Experiment 4. We trained
participants on the difference between single and double eyelids (a distinction that is appropriate for discriminating Asian faces but not for discriminating Caucasian faces). Our two control conditions (training on eyebrow darkness and training on face race) contained trainings that were not differentially more appropriate for discriminating either Asian or Caucasian faces. Pre-training, all participants exhibited the ORB in that Asians showed better discriminability ($d'$) for Asian faces than for Caucasian faces and Caucasians showed a higher $d'$ for Caucasian faces than for Asian faces. We predicted that the eyelid training would improve Caucasians’ recognition of Asian faces (and have no effect on Asians’ recognition of either Asian or Caucasian faces), but found that training participants to attend to this race-appropriate cue in a one-hour training task was insufficient to mitigate the ORB. Prior research has indicated that providing more extensive training and exposure to the appropriate cues (a combination of the contact theory and inappropriate cue utilization hypothesis) can successfully reduce the ORB (Hills & Lewis, 2006), and we will later discuss potential amendments to our training procedure that may have this desired effect.

Contact theory

The contact theory contends that the more interracial contact one has, the less he or she should exhibit the ORB (Brigham & Malpass, 1985). In all four experiments, we assessed the Asian participants as high-contact or low-contact (with reference to contact with Caucasians) based on the self-reported zip code for where each participant lived the longest in life. We gathered interracial contact data with an extensive questionnaire (see Figure 3.2), and this zip code correlated with the level of self-reported interracial contact for Asian participants. There was no measure that correlated with Caucasians’ self-
reported level of interracial contact with Asians, so we constrained our interracial contact analyses to Asian participants. Collecting zip code data for where each participant lived the longest in life is an important methodological distinction between the present set of studies and experiments currently in the literature. If researchers used demographics to infer level of interracial contact (e.g. Aronstam & Tyson, 1980; Platz & Hosch, 1988), they have typically done so using demographic data for the area where participants currently live at the time of the experiment rather than using demographic data from where participants lived the longest (and presumably could have the most influence, or lack of influence, from interracial contact). For example, many participants in our sample reported living in a different zip code area (than where they are currently) for the longest period in life, and these two zip codes have different population demographic statistics.

In Experiment 1, we found no difference between high- and low-contact Asians with respect to the descriptors they mentioned to describe Asian and Caucasian faces. Similarly in Experiment 4, we did not find any significant differences between high- and low-contact Asians on their recognition (measured in both $d'$ and $c$) of Asian or Caucasian faces. On the other hand, Experiment 2b revealed a significant interaction between level of interracial contact and test face race on false alarm rates. Recall that we discussed Experiment 2b as a better measure for interracial contact than Experiment 2a because of the more valid stimuli used in Experiment 2b. Asians high on interracial contact with Caucasians false alarms significantly less on Caucasian faces than did low-contact Asians, suggesting that these high-contact Asians may have had more experience with the appropriate cues necessary for Caucasian face discrimination.
We also found an interesting pattern with respect to interracial contact in Experiment 3. Asians high on interracial contact had an MDS algorithm that more closely resembled the Caucasian MDS algorithm than did the algorithm from the low-contact Asians. Specifically, Dimensions 3 (corresponding to eyebrow darkness and iris darkness) and 4 (corresponding to length of the chin, the ratio between forehead length to chin length, as well as the ratio between the length of the top half of the face to that of the bottom half, as divided at the base of the nose) from the high-contact Asian MDS solution correlated more strongly with Dimensions 3 and 6 from the Caucasian MDS solution than did Dimensions 3 and 4 from the low-contact Asian MDS solution. This trend supplements the result from Experiment 2 implying that not only do high-contact Asians show better recognition of Caucasian faces (in terms of fewer false alarms) than do low-contact Asians, but these high-contact Asians may have superior recognition because they are learning the appropriate cues.

What do our mixed findings regarding the contact theory tell us about the ORB? This pattern of inconsistent findings regarding interracial is not uncommon to the eyewitness and face recognition literatures, with many researchers finding support for a correlation between interracial contact and ORB (e.g. Chance et al., 1975; Brigham, Maass, Snyder, & Spaulding, 1982) while others do not (e.g. Malpass & Kravitz, 1969; Luce, 1974; Brigham & Barkowitz, 1978). If there were consistent findings in support of the contact theory, there would be no reason to see evidence in support of the inappropriate cue utilization hypothesis because the entire bias would be explainable by interracial contact alone. However, researchers have suggested (e.g. Malpass & Kravitz, 1969; Levin, 2000), and our results support the idea, that although contact is necessary to
mitigate the ORB, contact must occur along with learning the relevant cues appropriate for cross-race discrimination.

**Methodological limitations and future research directions**

We examined the inappropriate cue utilization theory and the contact hypothesis in all four experiments with Asian and Caucasian participants viewing Asian and Caucasian test faces. We were able to determine a measure that effectively divided the Asian participants into two groups of high- and low- interracial contact with Caucasians. Unfortunately, we were unable to establish a similar measure that divided Caucasians into differential contact groups. Consequently, we are able to draw conclusions about interracial contact only from the Asian participants.

Further, our measure of contact only allows us to infer the quantity of one’s interracial contact. It does not allow us to know what individuals are subsequently doing with this contact. In particular, how do we distinguish between those “high-contact” individuals who have learned more appropriate cues for cross-race face discrimination and those who have not? Research is needed to find a measure that makes this secondary distinction beyond categorizing someone as high- or low-contact.

We were able to identify iris color as a cue more important for discrimination between Caucasian faces (and more widely used by Caucasians) and presence/absence of an eyelid crease as a cue more important for discrimination between Asians faces (and more widely used by Asians). When we trained Caucasians to attend to eyelid creases, we found that this training did not decrease Caucasians’ expression of the ORB (as measured by $d'$). Other researchers (e.g. Hills & Lewis, 2006) have demonstrated that participants can show a decrease in ORB after training on features appropriate for other-
race discrimination. Perhaps if the training is conducted using faces of the same race as
the participants (as done by Hills and Lewis), participants would better learn the cues and
then transfer this knowledge to other-race faces. For example, future research should
train Caucasians to discriminate Caucasian faces on the basis of eyelid crease (though
this would include an unnatural manipulation of test faces) and then evaluate whether
Caucasians apply this knowledge to the discrimination of Asian faces. If Caucasians
demonstrate this transfer of information, this result would be consistent with the
inappropriate cue utilization hypothesis that people transfer cues used to discriminate
own-race faces to discriminate other-race faces.

In the present training experiment, we used the signal detection measures of $d'$
and $c$ to evaluate pre-training and post-training expression of ORB in participants.
Another means of assessing pre- versus post-training ORB is to use multidimensional
scaling. Comparing pre-training MDS solutions to post-training MDS solutions may
provide evidence for whether participants are perceptually using the cues on which they
are trained.

A follow-up from our MDS experiment comes from the finding that Dimension 4
from the Asian MDS solution correlates with Dimension 6 from the Caucasian MDS
solution (and refers to the length of the chin, the ratio between forehead length to chin
length, and the ratio between the length of the top half of the face to that of the bottom
half, as divided at the base of the nose). This dimension appears to be a more configural
dimension rather than a featural dimension. Recall we discussed (in the discussion of
Experiment 2b) the importance of configural information to face memory (Craik &
Tulving, 1975). Testing recognition memory for faces differing only on the basis of this
configural dimension may yield significantly different results (in terms of an interaction between participant race and test face race) than did our test of recognition memory (Experiment 2) for faces differing only on the basis of a featural dimension (iris color or the presence/absence of eyelid creases). Further evidence that configural cues are involved in memory comes from Hills and Lewis’s (2006) experiment training participants to discriminate faces with configural information rather than one specific feature. Training on configural cues effectively improved recognition memory for other-race faces and suggests that future research should involve examining the role of inappropriate [configural] cue utilization in the ORB.

**Concluding thoughts**

Results from the four experiments presented in this paper provide valuable contributions to the eyewitness and face recognition literatures in terms of: 1) methodologies used to investigate the inappropriate cue utilization and contact theories of the ORB and 2) the clarification of the inappropriate cue utilization theory and its interaction with the interracial contact hypothesis. The ORB is rarely investigated using Asians and Caucasians and is most often researched in terms of Blacks and Whites (e.g. Meissner & Brigham, 2001; Sporer, 2001). We used both perceptual-based (Experiment 1 with face descriptions and Experiment 3 with similarity ratings) and memory-based tasks (Experiment 2 with face recognition including faces with changed features and Experiment 4 comparing pre- to post-training face recognition). Finding significant results supporting cue utilization in the perceptual tasks suggests that individuals of different races perceptually encode different features and that this differential encoding influences recognition accuracy for same- and other-race faces. The fact that we did not
find support for differential cue utilization in the memory tasks is not necessarily evidence against the inappropriate cue utilization theory, but rather evidence that future research needs to investigate the influence of configural (rather than featural) cues on memory if in fact configural cues are more important than featural cues for memory (e.g. Craik & Tulving, 1975).

Furthermore, this series of four experiments examines the ORB in theory and in application. From a theoretical position, we examined a combination of the contact hypothesis in conjunction with the inappropriate cue utilization hypothesis as the most complete explanation for the mechanism driving the ORB. It is clear from Experiments 2 and 3 that individuals with more interracial contact may learn the appropriate cues for discriminating other-race faces and consequently demonstrate better recognition for other-race faces than do those individuals with less interracial contact. From an applied perspective, we were concerned with whether through training we were able to alter the extent to which people exhibit the ORB. Though the brief training did not produce the expected results, we believe based on the current literature that more extensive training could affect peoples’ recognition abilities such that we may see the bridge between theory and application.

The ambition of discussing the own-race bias in terms of learning and using race-appropriate cues for recognition is not to salvage any one theory, but rather to unify and further specify already-proposed theories. Although the ORB may not be driven entirely by inappropriate cue utilization or lack of interracial contact, it is important to understand how the theories complement one another to provide a more comprehensive picture of the own-race bias.
Experiment 1, Figure 1. Example test stimuli for Experiment 1.
Experiment 1, Figure 2. Example eyes with different iris colors. Clockwise from top left: hazel, grey, green, and blue.

Experiment 1, Figure 3. The eye on the left is an example eye with a “single eyelid.” The eye on the right is the same eye, but with a “double eyelid.”
Experiment 2, Figure 1. Example test stimuli for Experiment 2. The first Asian face in the top row and the first Caucasian face in the bottom row are the original faces. The second face in each row has the “opposite eyelid” from the original face. The third face in each row has the “opposite iris color” from the original face.
Experiment 3, Figure 1. Example test stimuli for Experiment 3. The top row contains one Asian male and one Asian female. The bottom row contains one Caucasian male and one Caucasian female.
Please respond to each of the following questions:

1. Gender:  M    F
2. Age:  __________
3. Race:  White  Black  Asian  Hispanic  Other (Please specify ______________)
4. Please indicate the zip-codes (ONLY) for the following addresses:
   a) current address:  __________
   b) birth address:  __________
   c) address where you have lived the longest:  __________

5. The following set of questions is intended to estimate the quantity and nature of your experience with people of the racial groups you viewed in the preceding pictures.

<table>
<thead>
<tr>
<th>Question</th>
<th>%</th>
<th>0%</th>
<th>1 – 25%</th>
<th>26 – 50%</th>
<th>51 – 75%</th>
<th>76 – 99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What proportion of children who attended your elementary school were WHITE?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. What proportion did you consider friends and playmates, or at least call by name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. What proportion of teachers at your elementary school were WHITE?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. What proportion of students at your high school were WHITE?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. What proportion of these did you know well enough to call by name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. What proportion were your close friends?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. What proportion of teachers at your high school were WHITE?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. What proportion could you call by name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. What proportion of people that you have worked with or for have been WHITE?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. What proportion could you call by name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Estimate the proportion of people you encounter in a typical day who are WHITE.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. What proportion of these could you consider friends or at least call by name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
13. What proportion of children who attended your elementary school were **Asian**?

14. What proportion did you consider friends and playmates, or at least call by name?

15. What proportion of teachers at your elementary school were **Asian**?

16. What proportion of students at your high school were **Asian**?

17. What proportion of these did you know well enough to call by name?

18. What proportion were your close friends?

19. What proportion of teachers at your high school were **Asian**?

20. What proportion could you call by name?

21. What proportion of people that you have worked with or for have been **Asian**?

22. What proportion could you call by name?

23. Estimate the proportion of people you encounter in a typical day who are **Asian**.

24. What proportion of these could you consider friends or at least call by name?

6. Think of the three (3) people in your life who you consider to be your closest friends (outside your family). Please provide the following information on each of those individuals:

**Individual #1**
- Gender: **M** **F**
- Age: __________
- Race: **W** **B** **A** **H** **O** (Please specify ____________________ )

How long have you known this individual?

_____ Less than 6 months  _____ 1 – 5 years
_____ 6 months – 1 year  _____ More than 5 years

How often do you see this individual?

_____ daily  _____ monthly
_____ weekly  _____ yearly
**Individual #2**

Gender: M F
Age: __________
Race: W B A H O (Please specify ____________________)

How long have you known this individual?

_____ Less than 6 months   _____ 1 – 5 years
_____ 6 months – 1 year   _____ More than 5 years

How often do you see this individual?

_____ daily     _____ monthly
_____ weekly     _____ yearly

**Individual #3**

Gender: M F
Age: __________
Race: W B A H O (Please specify ____________________)

How long have you known this individual?

_____ Less than 6 months   _____ 1 – 5 years
_____ 6 months – 1 year   _____ More than 5 years

How often do you see this individual?

_____ daily     _____ monthly
_____ weekly     _____ yearly

---

Experiment 3, Figure 2. Interracial Contact Questionnaire used in Experiment 3.
Experiment 3, Figure 3. Plots of the six dimensions revealed through multidimensional scaling in Experiment 3. These plots are based on Asian participants’ response data.
Experiment 3, Figure 4. Plots of the six dimensions revealed through multidimensional scaling in Experiment 3. These plots are based on Caucasian participants’ response data.
Experiment 4, Figure 1. Example test stimuli for Experiment 4.
Experiment 1, Table 1. Dimensions mentioned in Asian and Caucasian participants’ descriptions of Asian and Caucasian test faces

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Asian Participants</th>
<th>Caucasian Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race*</td>
<td>.87</td>
<td>.73</td>
</tr>
<tr>
<td>Gender</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>Age</td>
<td>.32</td>
<td>.28</td>
</tr>
<tr>
<td>Face length</td>
<td>.07</td>
<td>.12</td>
</tr>
<tr>
<td>Face width</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Face shape</td>
<td>.35</td>
<td>.29</td>
</tr>
<tr>
<td>Adiposity</td>
<td>.13</td>
<td>.02</td>
</tr>
<tr>
<td>Hair length</td>
<td>.57</td>
<td>.49</td>
</tr>
<tr>
<td>Hair color</td>
<td>.42</td>
<td>.38</td>
</tr>
<tr>
<td>Hair style</td>
<td>.53</td>
<td>.55</td>
</tr>
<tr>
<td>Forehead size</td>
<td>.06</td>
<td>.15</td>
</tr>
<tr>
<td>Eyebrow color*</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Eyebrow thickness</td>
<td>.19</td>
<td>.28</td>
</tr>
<tr>
<td>Eyebrow shape</td>
<td>.28</td>
<td>.24</td>
</tr>
<tr>
<td>Eye color* **</td>
<td>.13</td>
<td>.24</td>
</tr>
<tr>
<td>Eye shape</td>
<td>.41</td>
<td>.33</td>
</tr>
<tr>
<td>Inter-eye distance</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Eyelashes</td>
<td>.00</td>
<td>.07</td>
</tr>
</tbody>
</table>

(table continues)
Experiment 1, Table 1. (continued)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Asian Participants</th>
<th>Caucasian Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asian</td>
<td>Caucasian</td>
</tr>
<tr>
<td></td>
<td>.16</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>.45</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>.13</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.08</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>.41</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>.33</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>.06</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>7.23</td>
<td>6.87</td>
</tr>
</tbody>
</table>

Note. * significant main effect of participant race, ** significant interaction between participant race and test face race
Experiment 3, Table 1. Correlations between dimensions from Asian and Caucasian MDS algorithms

<table>
<thead>
<tr>
<th>Asian Dimensions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.995**</td>
<td>.032</td>
<td>.108</td>
<td>.006</td>
<td>.075</td>
<td>.028</td>
</tr>
<tr>
<td>2</td>
<td>.055</td>
<td>.983**</td>
<td>.068</td>
<td>.007</td>
<td>.024</td>
<td>.007</td>
</tr>
<tr>
<td>3</td>
<td>.014</td>
<td>.059</td>
<td>.892**</td>
<td>.252</td>
<td>.135</td>
<td>.169</td>
</tr>
<tr>
<td>4</td>
<td>.007</td>
<td>.072</td>
<td>.023</td>
<td>.480**</td>
<td>.081</td>
<td>.694**</td>
</tr>
<tr>
<td>5</td>
<td>.032</td>
<td>.050</td>
<td>.018</td>
<td>.398*</td>
<td>.557**</td>
<td>.276</td>
</tr>
<tr>
<td>6</td>
<td>.102</td>
<td>.032</td>
<td>.226</td>
<td>.521**</td>
<td>.553**</td>
<td>.206</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01
Experiment 3, Table 2. Correlations between dimensions from Asian (high and low interracial contact groups) and Caucasian MDS algorithms

<table>
<thead>
<tr>
<th>Asian Dimensions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 High</td>
<td>.996**</td>
<td>.028</td>
<td>.109</td>
<td>.005</td>
<td>.064</td>
<td>.033</td>
</tr>
<tr>
<td>Low</td>
<td>.993**</td>
<td>.064</td>
<td>.110</td>
<td>.025</td>
<td>.105</td>
<td>.015</td>
</tr>
<tr>
<td>2 High</td>
<td>.043</td>
<td>.982**</td>
<td>.085</td>
<td>.011</td>
<td>.027</td>
<td>.010</td>
</tr>
<tr>
<td>Low</td>
<td>.054</td>
<td>.983**</td>
<td>.041</td>
<td>.020</td>
<td>.019</td>
<td>.009</td>
</tr>
<tr>
<td>3 High</td>
<td>.019</td>
<td>.078</td>
<td>.923**</td>
<td>.179</td>
<td>.122</td>
<td>.111</td>
</tr>
<tr>
<td>Low</td>
<td>.014</td>
<td>.020</td>
<td>.809**</td>
<td>.335</td>
<td>.206</td>
<td>.144</td>
</tr>
<tr>
<td>4 High</td>
<td>.004</td>
<td>.071</td>
<td>.026</td>
<td>.428**</td>
<td>.030</td>
<td>.719**</td>
</tr>
<tr>
<td>Low</td>
<td>.021</td>
<td>.060</td>
<td>.001</td>
<td>.531**</td>
<td>.168</td>
<td>.582**</td>
</tr>
<tr>
<td>5 High</td>
<td>.018</td>
<td>.044</td>
<td>.058</td>
<td>.322*</td>
<td>.679**</td>
<td>.183</td>
</tr>
<tr>
<td>Low</td>
<td>.054</td>
<td>.025</td>
<td>.006</td>
<td>.262</td>
<td>.402**</td>
<td>.180</td>
</tr>
<tr>
<td>6 High</td>
<td>.134</td>
<td>.007</td>
<td>.132</td>
<td>.596**</td>
<td>.440**</td>
<td>.214</td>
</tr>
<tr>
<td>Low</td>
<td>.058</td>
<td>.001</td>
<td>.010</td>
<td>.117</td>
<td>.253</td>
<td>.330*</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01
Experiment 1, Graph 1. Caucasian participants mention iris color significantly more often than Asians when describing faces in Experiment 1. Further, all participants mention iris color significantly more when describing Caucasian faces than when describing Asian faces.

Experiment 1, Graph 2. Asian participants mention “single eyelids” and “double eyelids” significantly more often than Caucasians when describing faces in Experiment 1. Asians use this descriptor differentially more often for same-race Asian faces than for cross-race Caucasian faces.
Experiment 2, Graph 1. Participants in Experiment 2a false alarmed significantly more on Asian faces than on Caucasian faces in the “opposite eyelid” condition, and they false alarmed significantly more on Caucasian faces than on Asian faces in the “opposite iris” condition.

Experiment 2, Graph 2. Compared to the control condition, participants in Experiment 2a false alarmed significantly more on Asian faces in the “opposite eyelid” condition and less on Asian faces in the “opposite iris” condition. Participants also false alarmed significantly more on Caucasian faces in the “opposite iris” condition than compared to the control condition.
Experiment 2, Graph 3. Caucasian participants false alarmed more on Asian faces than on Caucasian faces, and Asian participants false alarmed more on Caucasian faces than on Asian faces in Experiment 2b. All participants false alarmed equally to Asian faces with the “opposite eyelid” and to Caucasian faces with the “opposite iris” color.
Experiment 4, Graph 1. Pre-training in Experiment 4, Asian participants set a higher decision criterion (c) for Caucasian faces than for Asian faces and Caucasians set a higher criterion for Asian faces than for Caucasian faces. This is evidence of the own-race bias.

Experiment 4, Graph 2. Across conditions in Experiment 4, Asian participants set a higher decision criterion (c) for Caucasian faces than for Asian faces and Caucasians set a higher criterion for Asian faces than for Caucasian faces. This is evidence of the own-race bias.
REFERENCES


