



Published in final edited form as:

Can J Exp Psychol. 2009 December ; 63(4): 253–262. doi:10.1037/a0015802.

Deficits in Other-Race Face Recognition: No Evidence for Encoding-Based Effects

Megan H. Papesh and Stephen D. Goldinger

Department of Psychology, Arizona State University.

Abstract

The other-race effect (ORE) in face recognition is typically observed in tasks which require long-term memory. Several studies, however, have found the effect early in face encoding (Lindsay, Jack, & Christian, 1991; Walker & Hewstone, 2006). In 6 experiments, with over 300 participants, we found no evidence that the recognition deficit associated with the ORE reflects deficits in immediate encoding. In Experiment 1, with a study-to-test retention interval of 4 min, participants were better able to recognise White faces, relative to Asian faces. Experiment 1 also validated the use of computer-generated faces in subsequent experiments. In Experiments 2 through 4, performance was virtually identical to Asian and White faces in match-to-sample, immediate recognition. In Experiment 5, decreasing target-foil similarity and disrupting the retention interval with trivia questions elicited a re-emergence of the ORE. Experiments 6A and 6B replicated this effect, and showed that memory for Asian faces was particularly susceptible to distraction; White faces were recognised equally well, regardless of trivia questions during the retention interval. The recognition deficit in the ORE apparently emerges from retention or retrieval deficits, not differences in immediate perceptual processing.

Keywords

face recognition; other-race effect; memory

In the past 50 years, research has consistently shown that people are extraordinary face processors, capable of recognising hundreds of faces across lengthy stretches of time (e.g., Bahrck, Bahrck, & Wittlinger, 1975). Despite such expertise, research also has consistently shown that people are relatively poor at learning and remembering faces from members of other racial groups (for a review, see Meissner & Brigham, 2001). This other-race effect (ORE; also, own-race bias, cross-race effect, etc.) has been well-documented under a variety of paradigms and is reliably observed across social and racial groups (e.g., Anthony, Copper, & Mullen, 1992; Bothwell, Brigham, & Malpass, 1989; Chance, Goldstein, & McBride, 1975; Ferguson, Rhodes, Lee, & Sriram, 2001; Ng & Lindsay, 1994). The effect has been documented in memory experiments with relatively short (e.g., 2 min) retention intervals (O'Toole, Deffenbacher, Valentin, & Abdi, 1994), and with retention intervals extending into days (Slone, Brigham, & Meissner, 2000). Such recognition differences do not apparently arise from concrete physiological differences amongst faces of different races (Goldstein, 1979), nor do they reflect negative racial attitudes (Slone et al., 2000; Swope, 1994). More interesting, race effects are not limited to long-term memory deficits; they also have immediate perceptual consequences, as researchers have found visual search asymmetries (Levin, 1996, 2000),

differences in perceived lightness (Levin & Banaji, 2006), and differences in perceived colour (Papesh & Goldinger, 2009), all as a function of depicted race. Race effects encompass myriad perceptual and memory effects, all of which hinge on an interaction between the race of the observer and the race (either subjective or objective) of the studied face.

The factor most commonly found to contribute to the presence and magnitude of the ORE is a person's relative degree of interracial contact. Contact hypotheses have been proposed in various forms and generally have been supported (e.g., Allport, 1954, as cited in Walker & Hewstone, 2006; Goldstein & Chance, 1980; Hancock & Rhodes, 2008, but see Brigham & Barkowitz, 1978, for an exception). All make similar claims, that recognition ability will be poor for members of one race who have had limited exposure to members of another race (Ng & Lindsay, 1994). This is an intuitively logical hypothesis. Gibson (1969), for instance, suggested that growing experience in any domain will lead to perceptual expertise, an increased ability to extract information necessary for accurate processing and recognition. This is also true of faces: As experience with a class of faces increases, so does the ability to quickly and accurately extract the information necessary for successful identification.

However, what information is extracted from faces, allowing such robust recognition? Researchers have consistently found that greater perceptual expertise corresponds to greater reliance on configural information from faces, relative to reliance on featural information (see Rakover, 2002, for an overview). Briefly, experts in any perceptual domain are more likely to process stimuli with emphasis on the Gestalt, the whole stimulus, rather than on serial consideration of its constituent parts (Diamond & Carey, 1986; Gauthier & Bukach, 2007; but see Robbins & McKone, 2007). For faces, configural information is contained within the spatial arrangement of the various features and their interrelations. When processing own-race faces, people have been found to rely most heavily on configural cues, whereas encoding other-race faces relies more heavily on featural processing, or the serial analysis of individual structures (Farah, Wilson, Tanaka, & Drain, 1998; Hancock & Rhodes, 2008; Michel, Caldara, & Rossion, 2006; Michel, Rossion, Han, Chung, & Caldara, 2006; Rhodes, Tan, Brake, & Taylor, 1989; Tanaka, Kiefer, & Bukach, 2004). It is commonly argued that such initial processing differences underlie the deficits seen in recognition memory tasks.

Typically, the ORE is expressed as better recognition memory for own-race faces, relative to other-race faces. This appears as both a decrease in sensitivity and as a bias shift, with increased false alarms to other-race faces. Many theories have been proposed to explain this bias, including multidimensional face-space frameworks (Byatt & Rhodes, 2004; Valentine, 1991; Valentine & Endo, 1992), contact hypotheses, feature-selection hypotheses (cf. Diamond & Carey, 1986; Hancock & Rhodes, 2008; Levin, 1996; Rakover, 2002), and notions of perceptual expertise (Lindsay, Jack, & Christian, 1991; Rakover, 2002), wherein members of one race are better able to extract the necessary individuating information from own-race faces, relative to other-race faces. Recent research, however, suggests that the ORE may not simply reflect long-term memory, but also differences in early perceptual processing. Race effects in immediate face perception are generally construed as differences in depths of processing: Other-race faces tend to be processed with emphasis on categorical information, whereas own-race faces are individuated more thoroughly, beginning at the earliest stages of processing (MacLin & Malpass, 2001; Ostrom, Carpenter, Sedikides, & Li, 1993; Sporer, 2001, but see Hayward, Rhodes, & Schwaninger, 2008, for evidence of an own-race advantage in both configural and component processing). Levin (1996, 2000) suggested that racial information is processed as a basic visual feature (cf. Triesman & Gelade, 1980), defined by its presence (as in other-race faces) or absence (as in own-race faces). According to Levin's theory, faces from different races can be thoroughly processed, with the same degree of individuation as own-race faces. Such complete appreciation, however, only occurs when people are motivated to do so. Oftentimes, people "stop encoding" beyond the initial processing of the basic race feature.

Such immediate perceptual effects of other-race face processing (e.g., Levin, 1996) are relevant, as they may predict that deficits will arise in immediate recognition memory, the focus of our investigation.

Lindsay et al. (1991) investigated immediate identification deficits in other-race face memory. Using faces with matched foils and a match-to-sample task, they found that own-race recognition was more accurate than other-race recognition, at least amongst White participants. The magnitude of their effect correlated negatively with the amount of interracial experience reported by their participants (i.e., people with greater experience were less susceptible to the ORE). Since that study, several other researchers have reported that faces are encoded differently based on race (Eberhardt, Dasgupta, & Banaszynski, 2003), such that memory deficits are immediately observable. Walker and Tanaka (2003) and Walker and Hewstone (2006) morphed faces along a biracial continuum and found evidence for an early perceptual effect, wherein own-race faces led to superior “same/different” discrimination. Such results are interesting, as they suggest that recognition memory deficits in the ORE may arise immediately, and are not the result of long-term memory processes, a hypothesis we examine in our experiments.

Overview

In the present study, we sought to replicate the match-to-sample results of Lindsay et al. (1991) with more stringently controlled stimuli. Lindsay et al. concluded their article by noting that other-race asymmetries in face recognition cannot be adequately explained unless both races are tested or, as they implied, researchers develop an objective measure of facial similarity. We have done the latter, by creating computer-generated faces, which we validated in Experiment 1. We then conducted a series match-to-sample experiments, wherein people saw a prime face, then tried to select it moments later in an AX task with itself (as a target) and a matched foil. By using computerized faces, we precisely controlled the physical similarity of targets (A) and foils (X) within and across races. In this methodology, immediate face encoding can be evaluated across depicted races, without confounds from overall similarity (or confusability) of the faces as a group. In every trial, people tried to discriminate a target face from a distorted version of itself.¹ Levels of distortion were equal across own- and other-race faces.

Using the computerized faces, we find that deficits in immediate face recognition are difficult (we refrain from saying “impossible”) to observe, and instead emerge in recognition tests following retention intervals during which immediate memory is disrupted. We achieved this by distracting participants during the retention interval, perhaps forcing them to rely on relatively long-term memory processes, changing the task from immediate to delayed recognition. Across experiments, we examined the time course of recognition deficits associated with the ORE by manipulating retention time and/or target-foil similarity. All experiments contrasted Asian and White face recognition. In Experiment 1, we tested whether participants experienced the classic ORE when given relatively lengthy encoding periods (3,200 ms per face), followed by a retention interval of several minutes. In Experiments 2 through 4, we varied target-foil similarity (and other factors) in match-to-sample tasks, seeking to determine whether the ORE can be elicited following briefer wait periods. Experiments 5 and 6 tested recognition accuracy for White and Asian faces, with reduced target-foil similarity and distracting questions presented during the retention intervals. Of the 368 participants included in the following experiments (including experiments reported in footnotes), 86% self-reported as White, 2% as Asian, and 12% as other (e.g., Black or Hispanic).

¹Note that this differs from the procedure used by Lindsay et al. (1991). In their study, target faces were discriminated from matched foils, which were different faces. In our study, target faces were discriminated from foils created by distorting the target face.

Experiment 1: Long-Term Recognition Memory Deficits

In Experiment 1, we sought to replicate the well-known ORE, wherein people excel at recognising faces from their own race while being relatively poor at recognising faces from another race (usually showing inflated false alarms). To this end, we conducted a discrete recognition memory experiment wherein participants studied a series of faces before making delayed recognition judgments. A secondary goal of conducting Experiment 1 was to validate a new set of stimuli: To provide tight experimental control for the match-to-sample tests, we developed a set of realistic face stimuli, which we could manipulate and present in lieu of photographic quality faces. As such, one of our aims in Experiment 1 was to ensure that our new stimuli were processed in the same general manner as actual faces (i.e., we wanted to verify that computerized faces evoke the ORE, like more naturalistic versions). Our computerized faces were all derived from actual photographs (sources are noted below). To validate our stimuli as acceptable proxy for real faces, we tested all participants with both the original photographs and their computerized versions. Both sets of stimuli were presented in separate, but otherwise identical, experiments, divided by a distractor task. We expected to find a recognition deficit for Asian faces, relative to recognition of the White faces, in both the real and computerized versions of the task.

Method

Participants—Thirty-six participants were recruited from the Arizona State University introductory psychology subject pool. They participated in exchange for partial course credit.

Stimulus materials—Original photographs of faces were gathered from the PAL database (Minear & Park, 2004), the Colour FERET database² (Phillips, Moon, Rizvi, & Rauss, 2000; Phillips, Wechsler, Huang, & Rauss, 1998), and the Ekman database (Ekman & Matsumoto, 1993). We selected sets of 80 Asian and 80 White faces, half male and half female, and generated computerized versions of each face using FaceGen Modeller software (Singular Inversions, Inc., 2004). This software equates the faces for size and allows modifications along 61 shape dimensions (e.g., brow–nose–chin ratio) and 36 texture dimensions (e.g., pale/flush) to more closely approximate the original face (see Figure 1). All original photographs were neutral in expression and lacked extraneous detail, such as glasses or facial hair. After generating the computerized faces, all the original photographs were edited with Microsoft Picture It software. Each face was cropped to remove most of the hair and resized to match the computerized versions. All faces, both photographic and computerized, were saved against a uniform background of grey (RGB = 140).

Design and procedure—Participants were asked to study 32 faces, 16 Asian and 16 White, for an upcoming recognition task. Stimulus type (real vs. computerized faces) was manipulated within-subjects, but between tasks (i.e., participants were not exposed to both types of face during a single task). Both memory tasks proceeded identically, as follows. Participants were instructed to press the spacebar on a “ready” screen to view each face, one at a time. After 3,200 ms had elapsed, the face disappeared and was replaced by the ready screen for the next trial. After all 32 faces had been studied, participants were given a 4 min break, during which time they were allowed to read, fill out puzzles, or simply sit quietly (most participants text messaged on their cell phones). Following the break, participants were shown 64 faces (half “old”) for a recognition task. “Old” and “new” responses were indicated via keypress and faces were presented individually in a response-terminated display. Feedback was not provided, and faces were presented in random order.

²Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the U.S. Department of Defense Counterdrug Technology Development Program Office.

Following the first memory task, all participants completed a “filler” task, which served as pilot data for another researcher in our lab. This was a lexical decision experiment using handwritten stimuli and lasted approximately 15 min. After the filler task was complete, participants completed the second face memory task, with the stimulus set not used previously. The presentation order of the stimulus sets was counterbalanced across participants.

Results and Discussion

Throughout all analyses reported in this article, a significance criterion of $\alpha = .05$ was adopted. Multiple comparisons were corrected for Type I error with Bonferonni adjustments. In Experiment 1, our analyses focused on signal-detection measures of sensitivity and bias. For each participant, we calculated d' and P_r , two measures of sensitivity, the former being the classic signal-detection measure and the latter from the “two-high threshold” model of recognition, representing the difference between hits (i.e., responding “old” when the face is old) and false alarms (i.e., responding “old” when the face is new; Feenan & Snodgrass, 1990; see Kleider & Goldinger, 2004). Also from the two-high threshold model, we calculated B_r , which is a measure of bias, defined as the probability of responding “old” despite uncertainty ($B_r = FA/(1 - P_r)$). B_r is centered around .5, with lower values reflecting a conservative bias.

The memory data were analysed in 10 repeated measures, oneway analyses of variances (ANOVAs), comparing race (Asian/ White) and contrasting stimulus (photo/computerized), for each of the five measurement types (hits, false alarms, d' , P_r , and B_r). In the interest of brevity, means for each measurement by stimulus type and race are shown in Table 1, and significant mean differences are noted.

As can be seen in the table, participants generally performed better in response to White faces, relative to Asian faces.³ Although hits did not differ between the races, false alarms were approximately 10% higher to Asian faces (both photographic and computerized). This was reflected in other measures of sensitivity and bias, indicating that participants were more liberal (B_r), less sensitive (d'), and had closer hit and false alarm rates (P_r) when judging Asian faces, relative to judging White faces. For the photographic faces, we observed numerical trends toward higher sensitivity for own-race faces, and a reliable difference in bias. For the computerized faces, we observed robust differences in sensitivity, without a difference in bias. These different statistical profiles reflect small differences in hit rates. Critically, the ORE was observed in both stimulus sets, suggesting that the ORE is reliably observed in response to photographic-quality faces and to computerized versions of those faces. Although the patterns of results across stimulus types were not identical, the data followed the same general trend, and the classic ORE is traditionally reflected in inflated false alarms.⁴ Thus, we used only the computerized faces in our subsequent experiments, allowing us to balance our stimuli while enacting various manipulations.

Experiments 2 Through 4: Match-to-Sample Tasks

Experiments 2 through 4 were designed to determine whether the recognition deficit associated with the ORE is the result of immediate, encoding-based differences, or other, postperceptual

³All of the results reported throughout this article remain unchanged by the inclusion of only White participants.

⁴In a subsequent replication using only the computerized faces (half of which were morphed distortions of the original faces used, selected from a pool of morphed faces created for Experiments 2 through 6) and a retention interval filled with 56 trivia questions (described later), rather than empty time, 28 participants demonstrated a similar pattern of results. Although hit rates were generally higher to Asian faces, $F(1, 28) = 6.46, p < .05, \eta_p^2 = .19$; with Asian accuracy at 76% and White accuracy at 68%, this was driven by a difference in bias (B_r), $F(1, 27) = 23.95, p < .05, \eta_p^2 = .47$ demonstrating that participants responded “old” more liberally to Asian faces. Sensitivity was also numerically higher for White faces (1.25) than Asian faces (0.94), although this difference was not statistically reliable.

factors. As was previously established, the ORE occurs when people study, and later try to recognise, other-race faces. Research also has shown immediate perceptual effects related to the ORE, such as differences in visual search (Levin, 1996, 2000), lightness perception (Levin & Banaji, 2006), and colour perception (Papesh & Goldinger, 2009). As noted earlier, Lindsay et al. (1991) found an other-race deficit in an immediate memory task, utilising a match-to-sample procedure. Participants in their experiment were exposed to masked prime faces for 120 ms each and then tried to make two-alternative forced-choice (2-AFC) decisions, choosing the targets over matched foils. The foils were paired to the targets based on the intuitions of a White researcher and a Black student (their study contrasted White and Black face recognition), leaving target-foil similarity subjectively controlled. Overall, in Experiments 2 through 4, we sought to replicate this effect with Asian and White faces, using objectively controlled stimuli. Simply stated, in each trial of Experiments 2 through 4, participants were presented with a prime face for study, and after a variable inter-stimulus interval (ISI), were shown the same face, now designated as the target, and a foil. Target-foil similarity was manipulated within and across experiments and, because of the manner in which we created the stimuli (see below), lightness and skin tone could not be used as reliable cues for recognition.

General Method

Participants—A total of 117 participants, recruited from the Arizona State University introductory psychology subject pool, participated in Experiments 2 through 4, in exchange for partial course credit. The precise numbers of participants per experiment are listed in Table 2. No participants were included in more than one experiment.

Stimulus materials—The computerized faces generated for Experiment 1 were used in all remaining experiments. In each trial of Experiments 2 through 4, participants saw a prime face, followed by that same face (as a target) and a matched foil: The matched foils were generated by systematically distorting each of the original 160 faces from Experiment 1 (in the interest of space, we will henceforth refer to computerized faces only as “faces”). Specifically, from each original face, we generated five distorted versions, yielding a total of 800 stimuli. The levels of distortion were increased, in a stepwise fashion, allowing us to make the discrimination task more or less challenging across trials. In creating distortions, we did not wish to consistently alter any single feature, which might become too salient over trials. Instead, we changed combinations of features, in three different pairs. Each distortion type involved simultaneously altering two facial features, eyes–nose, eyes–mouth, or mouth–nose (see Figure 2).

The modifications were made by concurrently altering the parameter values for the selected features on each individual face. Parameter values in the FaceGen software range between -10.00 and $+10.00$, and control the size of the features (e.g., larger/ smaller or wider/thinner) while maintaining realistic dimensions within the face contour. Because all faces were scanned from photographs of real people with intrinsic differences, each of the facial features began with different starting values across the faces. Nevertheless, from these starting points, all changes were quantitatively equal, with features modified in the same directions and degrees, relative to each original face. From the starting values, we modified the two features per distortion type by five different levels: two, three, four, five, and six points of difference. Half of the faces were positively distorted two of the five times, whereas the other half were positively distorted three of the five times. Each distortion type was used at least once per face, with the remaining two distortions assigned to random faces in a counterbalanced fashion, thereby allowing us to represent each distortion type equally throughout the experiments. Also, unlike Experiment 1, all faces were given a generic hairstyle to make them appear more realistic. All women received a chin-length hairstyle and all men received a short, standard

hairstyle. Asian faces were given black hair and White faces were given blonde hair (as in Figure 2), but otherwise the styles were identical.

Design and procedure—Participants were tested in groups of 2 to 8 on individual computers in the same room, separated by dividers. All experiments were programmed and controlled using E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA) and were presented on Gateway 15" CRT monitors with resolution of 1024×768 . Figure 3 provides a schematic outline of Experiments 2 through 6, although Experiments 5, 6A, and 6B will be discussed separately.

Experiments 2 through 4 were procedurally similar to one another, and are therefore described together: All participants were initially shown a black fixation cross against a white background for 1,500 ms, followed by a prime face for 1,500 ms. Once the prime face disappeared, a variable ISI was initiated, with duration manipulated within subjects (see Table 2). During the ISI, the screen was blank. At the termination of the ISI, participants were presented with a 2-AFC screen, and they indicated which face had been presented as the prime.⁵ Morphs and original faces were used equally often as targets and foils, and each participant completed 160 trials. Faces in the 2-AFC task were presented in a response-terminated display. Further details on each experiment, as well as the logical progression from one experiment to the next, are provided in tandem with the results.

Results and Discussion

Data from Experiments 2 through 4 were analysed in separate 2 (race) \times 3 (ISI) \times 5 (morph: two/three/four/five/six levels of distortion) ANOVAs. Accuracy means for each race in the separate experiments are presented in Table 3, organised by ISI (short, medium, long). Unless otherwise noted, there were no significant interactions in the data. Overall, Experiments 2 through 4 did not elicit any reliable other-race recognition deficits. Each experiment will be considered briefly, in turn.

In Experiment 2, we varied the retention intervals before the decision screen with ISIs of 250, 500, and 750 ms. Neither the main effect of race nor ISI was reliable, with no hints toward effects on recognition, both $F_s < 1.5$. In fact, the only variable to affect recognition was the distortion level of the faces (morph), $F(4, 11) = 23.84, p < .05, \eta_p^2 = .90$ (statistical power = 1.0). As is intuitively obvious, when foils were distorted to a greater degree, discrimination accuracy improved (see Figure 4).

In Experiment 3, we maintained the within-subjects manipulations of race and morph, and only altered the retention intervals between prime presentation and the 2-AFC screen, with ISIs of 250, 1,250, and 2,250 ms. Our reasoning (which also applied to the changes made in Experiment 4) was that extending the duration of the blank screen, which came after encoding but before test, might allow any latent ORE to gain strength. We hypothesised that, relative to the retention of own-race faces, participants would have greater difficulty maintaining accurate other-race face representations during longer retention intervals. As in Experiment 2, neither race nor ISI had a significant impact on accuracy, both $F_s < 3.0$. Again, the main effect of morph was significant, $F(4, 38) = 68.21, p < .05, \eta_p^2 = .88$ (statistical power = 1.0), as recognition accuracy increased when targets and foils differed more clearly.

⁵The design for Experiments 2 to 4 was motivated by an earlier experiment, wherein 48 participants were tested with masked primes of variable duration (250 to 750 ms) with only two-level morphs. Performance for each race ranged from 51% to 55% for the different durations, motivating us to drop the manipulation of prime duration and to use a standard presentation time.

Undaunted, we again manipulated ISI duration in Experiment 4, leaving every other variable unchanged from Experiments 2 and 3. We now used longer ISIs, of 1,250, 2,250, and 3,250 ms. Once again, only the main effect of morph was significant, $F(4, 56) = 71.54, p < .05, \eta_p^2 = .84$ (statistical power = 1.0). F values for both race and ISI were less than 1.5.

In contrast to Experiment 1, in which we observed an ORE after a retention interval of several minutes, we were repeatedly unable to observe race-based recognition deficits in Experiments 2 through 4, in which we varied ISI duration and target-foil similarity. Although it has been well-established that own- and other-race faces elicit differential processing (e.g., Lindsay et al., 1991; Walker & Hewstone, 2006; Walker & Tanaka, 2003), these processing differences do not seem to manifest themselves in an immediate recognition deficit. Despite prior findings, wherein immediate match-to-sample tasks elicited an ORE, our well-controlled stimuli were unable to elicit such an effect. Because of this, we are left to question whether the ORE in recognition memory is the result of encoding differences, as has been argued by Walker, Tanaka, Lindsay et al., and others, or whether they are an artefact of the inherent differences in the stimuli chosen for those experiments. That is, in Experiments 2 through 4, target-foil similarity had robust effects on performance. It is plausible that, in studies using natural faces, such similarity levels may be confounded with race. Indeed, discriminability differences are a key part of the ORE. Our experiments did not suffer this potential confound, as our stimuli were equated for differences between targets and foils across races and, as shown by the similar accuracy patterns across the distortion levels for each race, the distortions applied to the faces were psychologically equivalent across races. Recognition benefits increased, to equivalent degrees across both races, as morph levels increased.

Our results thus far suggest that, despite early perceptual differences that may arise in own- and other-race face processing, there is no apparent ORE in immediate face discrimination. Or, at any rate, we could not observe an effect, despite repeated attempts. To assess whether the effect can occur quickly, but just beyond face encoding, we conducted a final set of experiments, in which we found that target-foil dissimilarity and ISI interference are crucial for a recognition deficit to emerge.

Experiment 5: An Early, But Not Encoding-Based, Recognition Deficit

After repeatedly failing to demonstrate an ORE in immediate recognition tasks, we altered our experimental paradigm once more, to disrupt any potential processing during the retention interval. In Experiment 5, we only used one type of morph per experiment, a five-level distortion, reflecting the fact that accuracy was good (~86%), but not at ceiling, for five-level distortions in the previous three experiments.⁶ In addition, because our ISI manipulation resulted in impressively null effects on recognition accuracy, we deleted this variable. In place of a blank ISI screen, we inserted multiple-choice questions, again aiming to disrupt participants' processing. We hypothesised that distracting participants would induce greater forgetting of facial detail, which would allow the ORE to emerge in recognition.

Method

Participants—We recruited 44 students from the Arizona State University introductory psychology subject pool, none of whom had completed any of the previous experiments.

⁶Another reason we selected five-level morphs was because of the null results of another earlier experiment, conducted exactly as Experiment 5, but with four-level morphs. In that experiment, 66 participants demonstrated no reliable ORE, with White accuracy (61%) not significantly greater than Asian accuracy (59%), $t(1, 65) = -1.37, p > .05$.

Stimulus materials—We used a subset of same materials from the preceding experiments, specifically the original faces and their five-level morphed counterparts.

Procedure—The procedure was similar to that used in Experiments 2 through 4, but the ISI was now filled with multiple-choice academic and trivia questions (e.g., How many U.S. presidents have been assassinated while in office?), and therefore varied on a per-participant/per-trial basis. Briefly, participants viewed a prime face for 1,500 ms, which then disappeared and was replaced by the trivia question. After answering the trivia question, participants were presented with a 2-AFC screen containing the prime and a matched foil (see Figure 3). Responses were indicated via keypress.

Results and Discussion

The results were analysed in two paired-samples *t* tests, examining reaction time (RT) and accuracy. RTs to Asian faces (2,247 ms) and White faces (2,233 ms) were statistically equivalent, $t(1, 44) = -2.77, p > .05$. Although there were no race-based RT differences, we observed an ORE in accuracy, with better recognition for White faces (.71), relative to Asian faces (.66), $t(1, 44) = -2.77, p < .05$, Cohen's $d = -.49$ (Cohen, 1977). The results from Experiment 5 suggest that, for an ORE to emerge in recognition memory, the retention interval needs to be filled with a distraction, perhaps forcing degradation in the mental representation of other-race faces. Own-race faces, on the other hand, are apparently less affected by the disruption during the retention interval, demonstrating greater resistance to interference. In addition, target-foil similarity appears to be a major determinant in the emergence of the ORE. If the faces are too similar (e.g., level-two distortions), performance will be poor and nearly equivalent to faces from both races. If, however, the differences between target and foil faces are subtle, yet still discriminable (e.g., level-five distortions), participants will be able to make this discrimination more easily with own-race, versus other-race, faces.

Experiments 6A and 6B: Replications and Extensions

Although Experiment 5 confirmed the hypothesis that differences in memory for other-race faces are driven by processes beyond initial encoding, the effect was modest. As such, we conducted another two experiments, aimed to determine whether these effects would replicate, and whether the differences observed were truly the result of the disruptive trivia questions or if they would be observed without disruption. We hypothesised that participants would demonstrate poorer memory for other-race faces when their retention was disrupted by trivia, relative to an equivalent empty waiting period, potentially reflecting reliance on long-term memory for face matching.

Method

Participants—Thirty-four White participants were recruited from the Arizona State University introductory psychology subject pool and were randomly assigned to Experiment 6A (trivia) or Experiment 6B (no trivia). Five participants were dropped from analysis for consistent below-chance performance, leaving 17 participants in Experiment 6A and 12 in Experiment 6B.

Stimulus materials—Both experiments used the stimuli from Experiment 5 (original faces and their five-level morphed counterparts).

Procedure—The experiments were run between-subjects, following the trial-by-trial procedure of Experiment 5. Experiments 6A and 6B differed in one key regard: Whereas 6A was a direct replication of Experiment 5, Experiment 6B was a modified replication, replacing

the retention interval trivia questions with a 5,280-ms blank screen, which was the average RT to the same set of trivia questions by a group of six pilot test volunteers).

Results and Discussion

Accuracy data from both experiments were analysed in a 2 (Experiment: 6A/6B) \times 2 (race) mixed model, repeated measures ANOVA.⁷ Overall, there was no race effect, $F < 1.0$, and there was a marginal effect of experiment, $F(1, 27) = 3.81$, $p = .06$, $\eta_p^2 = .16$ (statistical power = .59). Participants who answered trivia questions before issuing 2-AFC decisions had an average accuracy of .73, whereas participants who did not answer trivia questions had an average accuracy of .79.

Planned comparisons examining the relationship between experiment and race indicate that differences in Experiment were only observed for the Asian faces, $F(1, 27) = 11.13$, $p < .05$, $\eta_p^2 = .29$ (statistical power = .90). When participants were required to answer trivia questions during the retention interval, they were an average of 10% less accurate to Asian faces, relative to when they did not have to answer questions (.71 vs. .81). Accuracy for White faces was statistically equal across experiments, $F < 1.0$. Experiments 6A and 6B provide evidence that the difference observed in Experiment 5 was reliable and the result of processing disruption during retention. If the differences resulted from encoding, they would have been observed in Experiment 6B, and the questions would not have had such a detrimental effect on performance in Experiment 6A.

General Discussion

In six experiments, we found evidence suggesting that the ORE in recognition memory is functionally distinct from OREs occurring early in perceptual processes (e.g., distortions in perceived lightness or colour, facilitated visual search). In Experiment 1, Asian faces yielded the typical recognition deficit, in comparison to White faces, following a retention interval of several minutes. A similar ORE in Experiment 1 was observed with computer-generated faces, validating their use in the 2-AFC tasks. In Experiments 2 through 4, we manipulated retention intervals and target-foil similarity (see Table 2 for a summary of manipulations by experiment): None of our manipulations had any impact on immediate face recognition in a match-to-sample task, which is known to result in an ORE (Lindsay et al., 1991). In Experiments 5 and 6A, we used target and foil faces that were more discriminable than the majority of faces used in the preceding experiments and we distracted participants with trivia questions during the retention interval. These manipulations caused the ORE to re-emerge in recognition accuracy. Experiment 6B highlighted the importance of distraction during the retention intervals in Experiments 5 and 6A: Comparing this experiment to the previous two indicated that only Asian faces were affected by the trivia questions; White face recognition accuracy remained stable. Of interest, using the same high-level distorted faces in Experiments 2 through 4 did not elicit an ORE. Although these experiments were procedurally similar to Experiments 5 and 6A, participants were not distracted during the retention intervals of those experiments, which may have prevented the emergence of the ORE. Overall, the results suggest that the ORE is the product of retention or retrieval, not a deficit in encoding.

Our results contrast with those of Lindsay et al. (1991), who found an ORE (amongst White participants) in a match-to-sample task contrasting Black and White face recognition. By their account, race-based asymmetries in face recognition could reflect intrinsic differences in the

⁷In the interest of brevity, we do not report RT analyses for these experiments, as the between- and within-experiment comparisons were null. There was, however, a numeric trend for participants in Experiment 6A to respond more slowly (2,232 ms), than participants in Experiment 6B (1,918 ms).

difficulty of the selected faces or real differences in the perceptual skills of participants. One potential limitation of the present research is that we were unable to represent White and Asian participants equally in our sample, a practise often used to study cross-race effects. This approach is typically used to ensure that, should an ORE be observed, it is not an artefact of inherent differences between the stimulus sets chosen for each race. Although we primarily sampled White participants, we were able to objectively equate the difficulty of the own-race and other-race trials, by using the same morph levels for the foil faces in each race. That is, each target face was presented with its own morphed counterpart, such that the difficulty of discrimination was quantitatively equal in Asian and White trials. Therefore, even if our sets of faces were not equally discriminable at the group level, it should not have affected our results in the match-to-sample tasks.

If our results reflected intrinsic differences in the difficulty of the chosen stimulus sets, we would have observed poor recognition abilities uniformly across our experiments. Instead, we observed relatively good performance to White faces and simultaneously poor performance to Asian faces, but only in tasks that tapped (relatively) long-term memory. Unlike Lindsay et al. (1991), we did not observe any clear race effects in our match-to-sample tasks, until Experiments 5 and 6A. Although our original motivations are not particularly relevant, we should note that our goal in this investigation was, in fact, to replicate the ORE in a test of immediate memory. In pursuit of this goal, we repeatedly tried to titrate performance, predicting the effect would emerge, but we could not make it happen. We observed this null effect in over 200 participants, leading us to conclude that the ORE in immediate recognition is elusive, at best.

Despite our inability to elicit immediate recognition deficits in other-race faces, we did observe a relatively early recognition deficit in Experiment 5, and its replication in Experiment 6A. Although these two experiments used a similar time-course as the experiments by Lindsay et al. (1991), we do not attribute the ORE to an immediate perceptual effect. Because we disrupted participants' rehearsal and maintenance during the retention interval, we may have forced them to rely on long-term representations when they made 2-AFC decisions. Because the similarity between the target and foil faces was equated between the Asian and White faces, we believe that our results reflect the inability to appreciate subtle changes in other-race faces, although such detail is preserved in the representation of own-race faces. This explanation fits nicely within the framework of multidimensional space (MDS) accounts of the ORE (Byatt & Rhodes, 2004; Valentine, 1991; Valentine & Endo, 1992). According to MDS accounts, people are ill-equipped to distinguish subtle differences amongst other-race faces, while being simultaneously expert at appreciating such differences in own-race faces. When faces are stored within a hypothetical "face space," they are arranged along dimensions useful for discriminating amongst stimuli. The points representing own-race faces are sparsely distributed throughout this space, reflecting the perceiver's enhanced ability to recognise subtle featural or configural changes. Points representing other-race faces, however, are more densely clustered, often toward the edges of the space, reflecting the perceiver's inability to accurately differentiate amongst them. Our results accord nicely with this explanation, as equivalent degrees of target-foil distortion across the races resulted in different levels of accuracy.

The ORE in recognition memory seems to emerge from long-term memory processing, not from differences in immediate encoding (although such differences clearly exist; Levin, 2000). The current experiments have shown that the ORE is near impossible to elicit in match-to-sample tasks when the similarity between target and foil faces is controlled across races. The effect only emerged when postperceptual processes, such as retention and rehearsal, were disrupted by a distraction task (in our experiments, we used lexical decision or trivia questions). Although our experiments do not represent an exhaustive set of immediate-memory tests of the ORE, we expended considerable effort in an attempt to elicit the effect, to no avail. The

effect seems to emerge in the process of remembering or retrieving faces, and future research should be guided toward determining the precise time course of this well-known effect.

Acknowledgments

Support provided by U.S. National Institutes of Health Grant R01–DC04535 to Stephen D. Goldinger.

References

- Anthony T, Copper C, Mullen B. Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin* 1992;18:296–301.
- Bahrack HP, Bahrack PO, Wittlinger RP. Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General* 1975;104:54–75.
- Bothwell RK, Brigham JC, Malpass RS. Cross-racial identification. *Personality and Social Psychology Bulletin* 1989;15:19–25.
- Brigham JC, Barkowitz P. Do “they all look alike?” The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology* 1978;23:306–318.
- Byatt G, Rhodes G. Identification of own-race and other-race faces: Implications for the representation of race in face-space. *Psychonomic Bulletin & Review* 2004;11:735–741. [PubMed: 15581126]
- Chance J, Goldstein AG, McBride L. Differential experience and recognition memory for faces. *The Journal of Social Psychology* 1975;97:243–253.
- Cohen, J. *Statistical power analysis for behavioral sciences*. New York: Academic; 1977. (Rev. ed.)
- Diamond R, Carey S. Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General* 1986;115:107–117. [PubMed: 2940312]
- Eberhardt JL, Dasgupta N, Banaszynski TL. Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin* 2003;29:360–370. [PubMed: 15273013]
- Ekman, P.; Matsumoto, D. *Japanese and Caucasian neutral faces (JACNeuF)*. San Francisco, CA: 1993. [Photographs on CD–Rom]
- Farah MJ, Wilson KD, Tanaka JN, Drain M. What is “special” about face perception? *Psychological Review* 1998;105:482–498. [PubMed: 9697428]
- Feenan K, Snodgrass JG. The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition* 1990;18:515–527.
- Ferguson GP, Rhodes D, Lee K, Sriram N. “They all look alike to me”: Prejudice and cross-race face recognition. *British Journal of Psychology* 2001;92:567–577. [PubMed: 11762861]
- Gauthier I, Bukach C. Should we reject the expertise hypothesis? *Cognition* 2007;103:322–330. [PubMed: 16780825]
- Gibson, EJ. *Principles of perceptual learning and development*. New York: Appleton; 1969.
- Goldstein AG. Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society* 1979;13:187–190.
- Goldstein AG, Chance JE. Memory for faces and schema theory. *The Journal of Psychology* 1980;105:47–59.
- Hancock KJ, Rhodes G. Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology* 2008;99:45–56. [PubMed: 17535471]
- Hayward WG, Rhodes G, Schwaninger A. An own-race advantage for components as well as configurations in face recognition. *Cognition* 2008;106:1017–1027. [PubMed: 17524388]
- Kleider HM, Goldinger SD. Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language* 2004;50:196–211.
- Levin DT. Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1996;22:1364–1382.
- Levin DT. Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General* 2000;129:559–574. [PubMed: 11142869]

- Levin DT, Banaji MR. Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General* 2006;135:501–512. [PubMed: 17087569]
- Lindsay DS, Jack PC Jr, Christian MA. Other-race face perception. *Journal of Applied Psychology* 1991;76:587–589. [PubMed: 1917773]
- MacLin OH, Malpass RS. Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law* 2001;7:98–118.
- Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law* 2001;7:3–35.
- Michel C, Caldara R, Rossion B. Same-race faces are perceived more holistically than other-race faces. *Visual Cognition* 2006;14:55–73.
- Michel C, Rossion B, Han J, Chung S-C, Caldara R. Holistic processing is finely tuned for faces of one's own race. *Psychological Science* 2006;17:608–615. [PubMed: 16866747]
- Minear M, Park DC. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers* 2004;36:630–633.
- Ng W, Lindsay RCL. Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology* 1994;25:217–232.
- O'Toole AJ, Deffenbacher KA, Valentin D, Abdi H. Structural aspects of face recognition and the other-race effect. *Memory & Cognition* 1994;22:208–224.
- Ostrom TM, Carpenter SL, Sedikides C, Li F. Differential processing of in-group and out-group information. *Journal of Personality and Social Psychology* 1993;64:21–34.
- Papesh MH, Goldinger SD. Visual search for racially ambiguous faces: A test of the race-feature hypothesis. 2008 Manuscript submitted for publication.
- Phillips PJ, Moon H, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans Pattern Analysis and Machine Intelligence* 2000;22:1090–1104.
- Phillips PJ, Wechsler H, Huang J, Rauss P. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing Journal* 1998;16:295–306.
- Rakover SS. Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology* 2002;93:1–30. [PubMed: 11839099]
- Rhodes G, Tan S, Brake S, Taylor K. Expertise and configural coding in face recognition: Erratum. *British Journal of Psychology* 1989;80:526.
- Robbins R, McKone E. No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition* 2007;103:34–79. [PubMed: 16616910]
- Singular Inversions, Inc. FaceGenModeller (Version 3.1.4) [Computer software]. Toronto, ON, Canada: 2004. Available: <http://www.FaceGen.com>
- Slone A, Brigham J, Meissner C. Social and cognitive factors affecting the own-race bias in Whites. *Basic and Applied Social Psychology* 2000;22:71–84.
- Sporer SL. Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law* 2001;7:36–97.
- Swope, TM. Social experience, illusory correlation and facial recognition ability. Florida State University; 1994. Unpublished master's thesis
- Tanaka JW, Kiefer M, Bukach CM. A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition* 2004;93:B1–B9. [PubMed: 15110726]
- Triesman AM, Gelade G. A feature-integration theory of attention. *Cognitive Psychology* 1980;12:97–136. [PubMed: 7351125]
- Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology* 1991;43A:161–204. [PubMed: 1866456]
- Valentine T, Endo M. Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology* 1992;44A:671–703. [PubMed: 1615169]
- Walker PM, Hewstone M. A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology* 2006;20:461–475.
- Walker PM, Tanaka JW. An encoding advantage for own-race versus other-race faces. *Perception* 2003;32:1117–1125. [PubMed: 14651324]

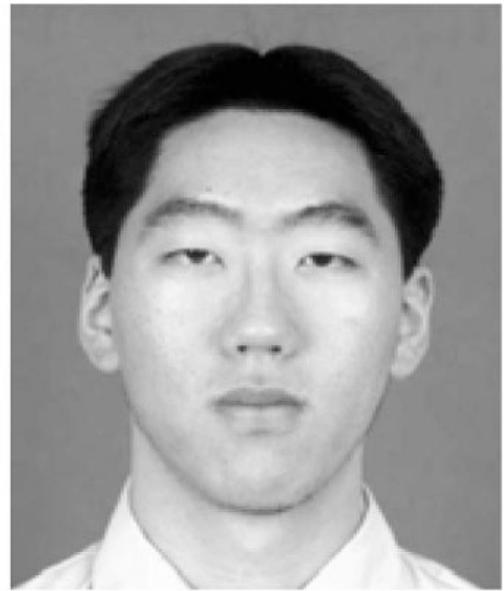


Figure 1.
Examples of two original photographs and their computerized copies.













Original	Level 2	Level 3	Level 4	Level 5	Level 6
					
					
	Nose/Mouth	Eyes/Mouth	Nose/Mouth	Eyes/Nose	Eyes/Mouth

Figure 2.

Example of face stimuli used in Experiments 2 through 6B. The distortion level is noted above the faces and the distortion type is noted below. Note that each distortion type was not limited to the distortion levels depicted in the example. Each distortion type was created at each level. The apparent stripe across the forehead of the White face is a printing artefact and was not visible on the computer screen for participants.

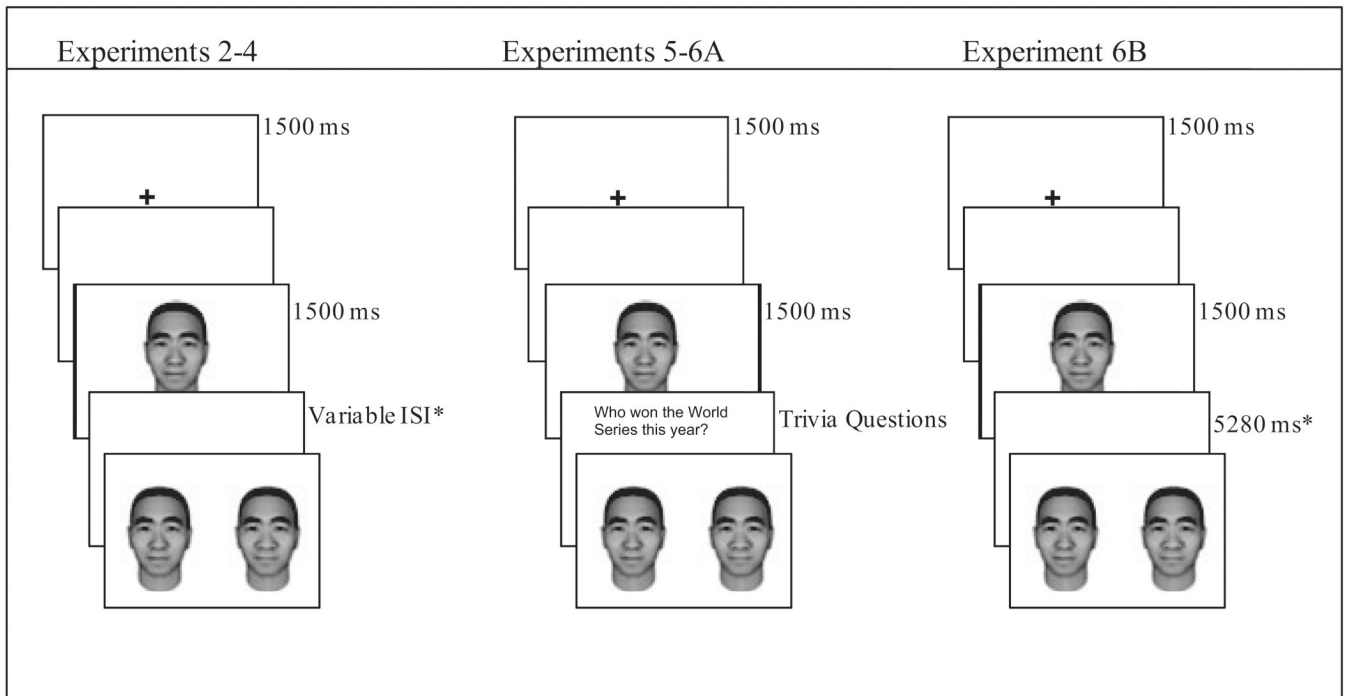


Figure 3. A general outline of trials in Experiments 2 through 6B. Unless otherwise labelled, blank screens were presented for 40 ms. *See text for details. ISI = interstimulus interval.

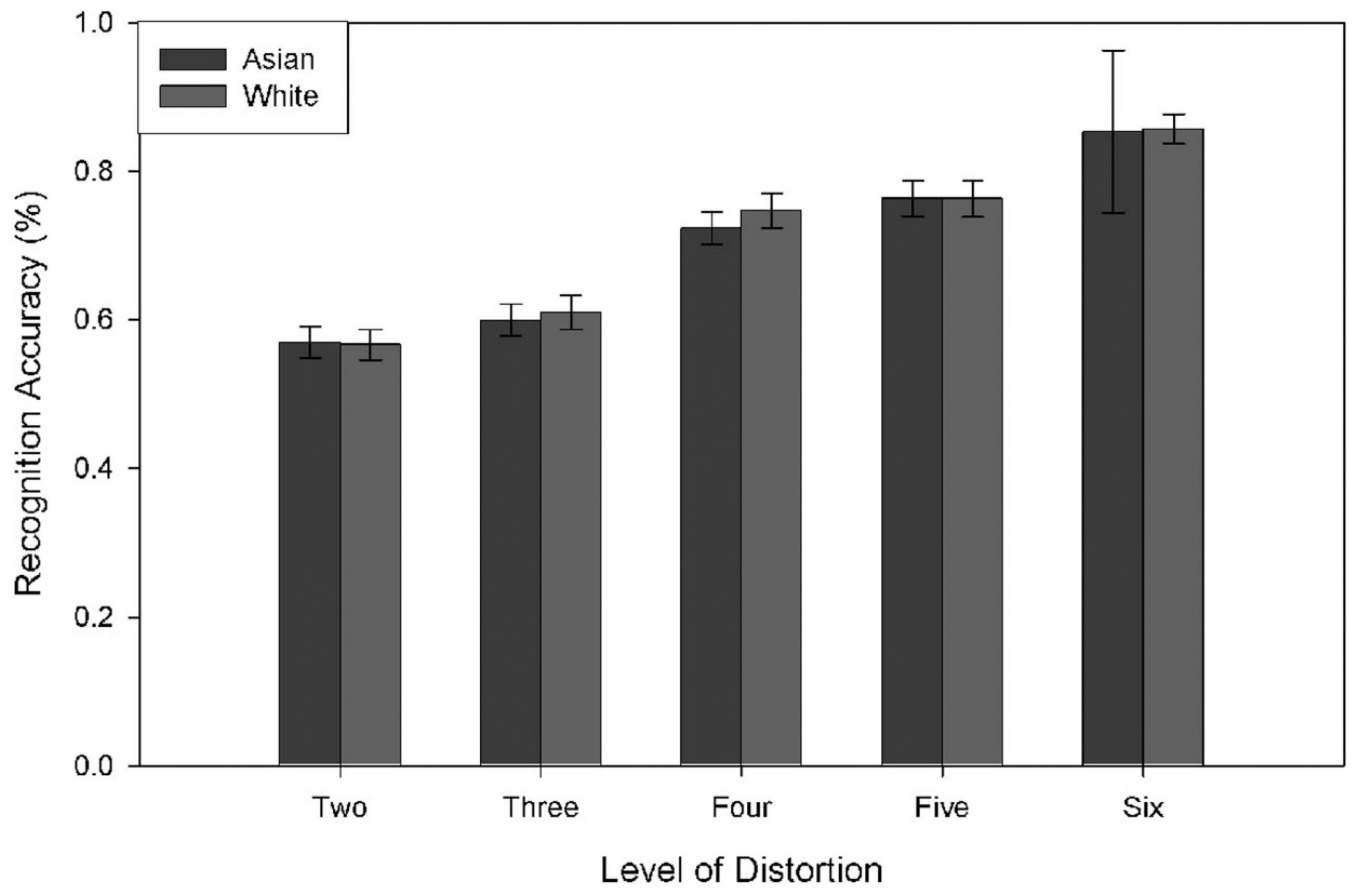


Figure 4. Accuracy means for each distortion-level difference between primes and foils for each race, collapsed across Experiments 2 through 4. Error bars represent standard errors.

Table 1

Recognition Results and Signal Detection Indexes From Experiment 1

Stimulus type	Race	Hit rate	FA rate	d'	P_r	B_r
Photographic	Asian	.75	.37* (.16)	1.14	.38	.61* (.16)
	White	.71	.28* (.16)	1.24	.43	.48* (.16)
Computerized	Asian	.73	.51* (.21)	0.73* (.16)	.22* (.25)	.66
	White	.75	.40* (.21)	1.10* (.16)	.35* (.25)	.64

Note. Measures of effect size (η_p^2) are in parentheses for the significant effects. FA = false alarm.

* $p < .05$.

Table 2

Summary of Main Methodological Features in Experiments 2 Through 6B

Experiment	ISI (ms)	Morphs	<i>n</i>
2	250/750/1,250	All five levels	15
3	250/1,250/2,250	All five levels	42
4	1,250/2,250/3,250	All five levels	60
5	Varied: Trivia	Level 5	44
6A	Varied: Trivia	Level 5	17
6B	5,280 ^a	Level 5	12

Note. ISI = interstimulus interval.

^aSee text for details.

Table 3

Accuracy Means in Experiments 2 Through 4, Organized by ISI

Experiment	Asian			White		
	Short	Medium	Long	Short	Medium	Long
2	.71	.73	.71	.73	.73	.74
3	.70	.69	.69	.72	.71	.70
4	.71	.69	.69	.69	.69	.69

Note. ISI = interstimulus interval.