Journal of Experimental Psychology: General
2000, Vol. 129, No. 4, 559-574

Copyright 2000 by the American Psychological Association, Inc.
0096-3445/00/$5.00 DOI: 10.1037/0096-3445.129.4.559

Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit

Daniel T. Levin
Kent State University

One of the most familiar empirical phenomena associated with face recognition is the cross-race (CR) recognition deficit whereby people have difficulty recognizing members of a race different from their own. Most researchers assume that the CR deficit is caused by failure to generalize perceptual encoding expertise from same-race (SR) faces to CR faces. However, this explanation ignores critical differences in the social cognitions and feature coding priorities associated with SR and CR faces. On the basis of data from visual search and perceptual discrimination tasks, it appears that the deficit occurs because people emphasize visual information specifying race at the expense of individuating information when recognizing CR faces. In particular, it is possible to observe a paradoxical improvement in both detection and perceptual discrimination accuracy for CR faces that is limited to those who recognize them poorly. These findings support a new explanation for the CR recognition deficit based on feature coding differences between CR and SR faces, and appear incompatible with similarity-based models of face categories.

In some cases, the simplest phenomena are the most mysterious. Perhaps this is because a correspondingly simple explanation is expected to suffice, and when it doesn't the problem becomes doubly opaque. In this article, I explore the mystery surrounding the cross-race (CR) recognition deficit by using visual search and perceptual discrimination tasks. Although it has long been known that people have difficulty recognizing members of other races (i.e., Feingold, 1914; Malpass & Kravitz, 1969), psychologists have failed to produce a definitive explanation for this phenomenon (Teitelbaum & Geiselman, 1997). Perhaps this is because the simple explanation that our lack of contact with CR individuals makes us less expert at coding their faces has not received consistent empirical support (Ng & Lindsay, 1994). Another version of the expertise hypothesis is based on the assumption that CR faces are more psychologically similar as a group than same race (SR) faces and are therefore more confusable (Valentine, 1991; Valentine & Endo, 1992). I argue that both of these explanations are inadequate because they make simplifying assumptions about face categories. The most important simplification is the assumption that object categories such as groups of faces are best understood by modeling abstract similarity relationships among category exemplars. This deemphasizes the process of selecting the features that will be used to make the similarity comparison.

This kind of simplification causes exemplar models of face categories to exclude explicit consideration of the complex web of social cognition that shapes these categories. These cognitions are the foundation for a process that differentially selects race-specifying information in CR and SR faces. In particular, CR persons are coded with more emphasis on category-related information (e.g., stereotypes that apply to an entire group) and less individuating information than are SR persons (Anthony, Cooper, & Mullen, 1992; Ostrom, 1993). This may lead to an asymmetric feature-selection process that causes race-specifying information to be coded as a visual feature in CR faces and not in SR faces (Levin, 1996). In this article, I propose a new explanation for the CR recognition deficit which assumes that selecting race-specifying information in CR faces reduces the amount of individuating information available and therefore reduces recognition accuracy in individuation tasks. Thus, the CR recognition deficit is not caused by failure to generalize perceptual expertise to CR faces; rather, it occurs because the information people select in CR faces is optimal for classification and not recognition. I also show that this feature-selection process has implications for visual search and perceptual discrimination tasks that are fundamentally incompatible both with perceptual expertise explanations of the CR deficit and with similarity-based models of face categories.

The Mystery Surrounding the CR Recognition Deficit

It has long been assumed that recognizing members of a race different from one's own is more difficult than recognizing SR faces (see for example Feingold, 1914). The CR recognition deficit has been confirmed repeatedly in a wide variety of participants (for reviews, see Anthony et al., 1992; Bothwell, Brigham, & Malpass,

Portions of this work were presented at the meeting of the Association for Research in Vision and Ophthalmology conference, Ft. Lauderdale, Florida, April 1996, and were reported as part of Daniel T. Levin's doctoral dissertation submitted to Cornell University. Preparation of this article was supported by National Institute of Mental Health Grant 1-R03-MH60137-01.

I would like to thank Laura Bogart, David Dunning, Frank Keil, Beena Khurana, Bill Merriman, Julia Noland, and Dan Simons for reading and commenting on previous versions of this article.

Correspondence concerning this article should be addressed to Daniel T. Levin, Department of Psychology, Kent State University, Kent, Ohio 44242-0001. Electronic mail may be sent to dlevin@kent.edu.

1989; Shapiro & Penrod, 1986).[1] Yet, for all the research on this topic, the basic cause of the deficit remains in many ways a mystery (Ng & Lindsay, 1994). The major problem seems to be the inconsistent empirical success of the *contact hypothesis,* which states that increased contact with CR individuals should increase recognition accuracy. For instance, the prediction that living among members of another race will reduce the CR recognition deficit has been supported in some experiments (Carroo, 1986; Chiroro & Valentine, 1995; Cross, Cross, & Daly, 1971) but not in others (Ng & Lindsay, 1994). The findings are also mixed for other measures of contact. Attending integrated or predominantly CR schools has in some cases been shown to reduce the deficit (Feinman & Entwhistle, 1976), whereas in another case it had no effect (Malpass & Kravitz, 1969) or has even resulted in an increased CR recognition deficit (Lavarkis, Buri, & Mayzner, 1976). Finally, Brigham and Barkowitz (1978) found no relationship between self-reported CR contact and the recognition deficit, whereas Li, Dunning, and Malpass (1994) found that White basketball fans show no CR recognition deficit in recognizing Black faces.

The failure of the contact hypothesis is mysterious because at some level it *has* to be true unless one posits an innate inability to accurately code CR faces. Perhaps these findings suggest that only specific kinds of contact will improve CR recognition. For example, a number of authors have argued that contact must involve individuation of CR faces (Shepard, 1981; Valentine, Chiroro, & Dixon, 1995) and, further, that many interactions that are characteristic of living among members of another race do not require that individuals be remembered (Li et al., 1994). Therefore, measures of contact that do not require individuation could reveal no effect, but contact situations requiring individuation may reduce the deficit. Accordingly, White basketball fans tend to individuate a large number of Black players and therefore might be expected to show no CR deficit.

It appears, therefore, that mere contact is not sufficient to eliminate the CR recognition deficit. Instead, the perceiver must process CR faces with the goal of individuation. This bears an interesting relationship to research on the social cognitions associated with group membership, which provides a principled explanation of coding differences associated with CR and SR faces. In the next section, these coding differences are reviewed and related to a visual search task that confirms the prediction that visual information specifying group membership is coded in CR but not in SR faces. This feature-coding asymmetry will be used as the basis for a new explanation of the CR recognition deficit.

Face Categories and Visual Features: A Search Asymmetry for Race

A long tradition of research in social cognition reinforces the assumption that CR individuals have a more salient category-level identity than SR individuals (Sedikides & Ostrom, 1993). This research has historically been guided by the need to understand the formation and operation of racial and ethnic stereotypes that are assumed to reflect basic categorization processes (Taylor, Fiske, Etcoff, & Ruderman, 1978). It is clear from this research that simple models of differential familiarity between groups cannot adequately explain the rich variety of cognitions that are associated with them (Sedikides & Ostrom, 1993). Instead, it has been necessary to account for a variety of contextual factors that affect the way that CR and SR categories are coded. For present purposes, the most important of these is that members of a given race may belong to a group that is saliently different from one's own and therefore constitute an "out-group" that contrasts with members of one's own group (an "in-group"). In a test of the implications of the in-group/out-group contrast, Linville, Fischer, and Salovey (1989) found that participants considered members of out-groups to be more homogeneous than members of in-groups and that they believed group stereotypes to be more descriptive of out-group members. In Linville et al.'s studies, in-groups and out-groups were defined by age and nationality, whereas other studies have focused on groupings such as sorority and fraternity membership (Bartsch & Judd, 1993), and have even created novel groupings (Simon & Brown, 1987). Accordingly, across a wide variety of situations, cognitions about out-group members emphasize category-level information, whereas cognitions about in-group members emphasize cognitions about individuals.

On the basis of findings like these, I have argued (Levin, 1996) that visual processing of faces might reflect these cognitive contrasts. In particular, I proposed that the contrast between CR and SR faces is coded as a feature-present/feature-absent relationship in which people code race-specifying information in CR faces as a feature but code SR faces as having no such feature. This coding relationship was tested with a visual search task based on those used by Triesman and Gormican (1988), who showed that visual search for a feature-positive target among feature-negative distractors is faster than the reverse (i.e., searching for a feature-negative target among feature-positive distractors). For example, a straight line might be considered a standard stimulus, whereas a tilted line would be considered a feature-positive variant of this standard because it is coded by the visual system as a straight line plus a "tilt" feature. This effect, called a *search asymmetry,* was assumed to occur because locating a feature-positive target is easy—it stands out from the feature-negative distractors. Search for a feature-negative target among feature-positive distractors is difficult because the distractors effectively obscure the target in noise. This reasoning can easily be applied to CR and SR faces. Instead of a straight line, the feature-negative norm is an SR face, and

[1] Here and throughout I will refer to recognition of faces from an unfamiliar group as *cross-race recognition* to remain consistent with existing terminology, despite the fact that the experiments include about 10–15% Asian participants for whom both the White and Black faces are cross-race. Inclusion of Asian participants should have minimal impact on the results because, similar to White participants, they tend to recognize Black faces less accurately than they do White faces (Teitelbaum & Geiselman, 1997).

To ensure that this strategy did not impact on essential conclusions, I reanalyzed data from all four experiments with White participants only. In Experiment 1, the critical Group × Task × Display Size interaction remained significant, $F(2, 32) = 3.550$, $MSE = 11{,}156$, $p = .0405$. In Experiment 2, the Task × Display Size interaction was still significant, $F(2, 20) = 4.462$, $MSE = 14{,}866$, $p = .0250$. In Experiment 3, the effect of continuum end remained significant, $F(1, 6) = 10.559$, $MSE = 0.011$, $p = .0173$. In Experiment 4, the Group × Continuum End effect narrowly missed conventional levels of significance, $F(1, 15) = 4.173$, $MSE = 0.007$, $p = .0590$. In this experiment the continuum end effect was still significant for the deficit group, $F(1, 15) = 9.788$, $MSE = 0.007$, $p = .007$, and did not approach significance for the no-deficit group ($F < 1$).

instead of a line-plus-tilt, the feature-positive stimulus is a CR face (i.e., a norm face + CR race-feature). In this case, the visual search prediction is clear. For White participants, searching for the feature-positive Black face among feature-negative White faces should be easier than the reverse. The single Black face should "stand out" among White faces more than the single White face stands out from a field of Black faces.

Levin (1996) confirmed this prediction using a standard visual search task. White participants searched more quickly for a Black face among White faces (52 ms/item in positive trials) than the reverse (75 ms/item). In addition, two groups of Black participants were tested: African Americans, and African nationals. The African national participants showed no asymmetry at all, and the African American participants showed results intermediate between the Whites and African nationals. Thus, this finding cannot be accounted for by simple perceptual differences between Black and White faces and must be understood in the context of the specific learning histories of these different groups. It is important to note here that for members of minority groups, the tendency to code out-groups at the category level is probably moderated by the tendency to code in-group members at the category level as well (Bartsch & Judd, 1993; Simon & Brown, 1987). This could explain why the search asymmetry did not completely reverse for the African American participants who, as minorities in America, might not be expected to show the exact converse effect of White, majority subjects.

The research reviewed above suggests a feature coding asymmetry whereby CR faces have race-specifying information that is lacking in SR faces, which can be understood as the direct result of social cognitions associated with perception of persons. This section began with the suggestion that the contact hypothesis may need to specify contact through individuation. This orientation is probably affected by in-group/out-group cognitions, which can include a variety of complex contextual factors. Therefore, the resultant feature-coding asymmetry would not be caused by simple differences in perceptual expertise or familiarity between the races. It seems better conceptualized in terms of a feature-selection process in which the visual features used for object recognition are chosen by the ecological context of recognizing socially meaningful objects. It therefore seems plausible to argue that the CR recognition deficit itself is not the result of simple differences in expertise, but rather is mediated by social cognitions. In the next section, the feature-selection explanation for the CR recognition deficit is proposed.

The Feature-Selection Explanation of the CR Recognition Deficit

According to the feature-selection hypothesis, the CR recognition deficit occurs because people code race-specifying features at the expense of individuating information. Although many current researchers tend to discount this possibility (see, e.g., Chiroro & Valentine, 1995), the first articles on face recognition explicitly suggested that category-level cognition can interfere with individuation. For example, Malpass and Kravitz (1969) proposed that

> if physiognomic discriminations among persons of other race are frequently made, within-group discriminative cue utilization habits will be formed. If contrasts of persons of other race from own race are predominant, between-group cue utilization habits will be formed. The latter may serve to decrease the use or acquisition of the former. (p. 333)

Shepard (1981) suggested a similar possibility, arguing that overuse of the CR category can cause the deficit. More recently, Anthony et al. (1992) argued that SR faces are processed with an "exemplar" strategy that emphasizes individuals, whereas CR faces are processed with a "prototype" strategy that explicitly relates individuals to the prototype of their group category. In all of these cases, the deficit is not caused by differences in perceptual expertise between groups. It is instead based on differences in the kinds of information that is coded between the races. In CR faces, group-specifying information is coded at the expense of individuating information. Because most recognition tasks depend on individuating information, performance is poor when CR faces must be differentiated. This shortcoming would be similar to children's overuse of peripheralia such as glasses and hats when recognizing faces (Carey & Diamond, 1977; Flin, 1985). Instead of focusing on decisive individuating information, they also include less decisive information that can reduce recognition accuracy.

One reason this kind of hypothesis has not been seriously considered is that most researchers have concluded that, despite failures of the contact hypothesis, the CR recognition deficit is based in a difference in coding expertise or skill between the races. For example, according to Lindsay, Jack, and Christian (1991), "there . . . appears to be wide-spread agreement that the locus of the effect is encoding and, more specifically, that the effect reflects race-related differences in perceptual expertise" (p. 587). Although an encoding expertise explanation for the CR recognition deficit is highly intuitive, currently available data suggest that it might be inadequate. Several studies show that it is possible to eliminate the deficit with a relatively small amount of training. Goldstein and Chance (1985) eliminated the deficit with about 3 hr of individuation training spread over a few weeks. Further, by the time their "no training" control group (who completed a series of recognition tests without the intensive training) completed the experiment, they too showed no recognition deficit. In fact, several earlier studies had shown that the deficit could be reduced, if only temporarily, in a single hour-long session (Elliott, Wills, & Goldstein, 1973; Lavarkis et al., 1976; Malpass, Lavigueur, & Weldon, 1973). If the CR recognition deficit is a skill deficit built up through long experience, then one would not expect an hour's worth of training to eliminate it. At this point, however, it is possible to retain an encoding expertise hypothesis by assuming that the perceptual expertise used in SR face processing needs only the slightest push to generalize to CR faces.

More data are therefore necessary to test the feature-selection hypothesis as an explanation for the CR recognition deficit. First, in Experiment 1, the search asymmetry favoring CR faces is related to the CR recognition deficit by separating participants who do not show a recognition deficit from those who do. The prediction is that individuals who show no recognition deficit will also show no search asymmetry or a reduced one. In Experiment 2, the feature-selection hypothesis is separated from a common expertise-based hypothesis by having participants search for inverted faces. The logic behind this manipulation will be more fully explained in the introduction to that experiment. Finally, Experiments 3–4 test for increased sensitivity to race-based variation in

CR faces in a discrimination task. If the contrast between CR and SR face reflects differential coding of race-specifying information, then it should be possible to observe a paradoxical increase in discrimination accuracy among CR faces if the task depends on this information. In Experiments 3 and 4, this hypothesis is tested with a perceptual discrimination paradigm to confirm that participants are in some cases better able to discriminate among different versions of CR faces.

Experiment 1

This experiment was undertaken to replicate the basic search asymmetry observed in Levin (1996), and to test for a relationship between the search advantage for CR faces and the CR recognition deficit. To test this hypothesis, participants completed a face recognition test after doing the search task. By means of this test, a group of participants who did not show the typical CR recognition deficit was isolated. It was predicted that these participants would also show a reduced search asymmetry, and that those who did show this deficit would show a strong search asymmetry favoring detection of Black faces.

In addition, the perceptual identification task I used earlier (Levin, 1996) was included in this experiment. In this task, participants viewed a randomly alternating series of individual Black and White faces. In one block of trials, participants responded by hitting a key when they saw the Black face and did nothing when they saw the White face. In the other block of trials the roles were reversed. This task is meant to rule out an alternative explanation for the asymmetry. It is possible that searching for the Black target is faster than searching for the White target because participants can quickly identify and reject White distracters when they scan through them in search of the Black face. This explanation for the asymmetry predicts that White faces should be detected more quickly than Black faces when no distracters are present. I found the reverse (Levin, 1996), and the prediction here is the same. In addition, participants who show no CR recognition deficit should also show no perceptual identification advantage for CR faces.

Method

Participants. Twenty-five participants completed Experiment 1. For all experiments in this report, participants who made more than 5% response errors were eliminated from the analysis. In the present experiment, 4 participants failed to meet this criterion, leaving a total of 21 participants in the analysis. The average error rate for those remaining was 1.9%. Of these, 17 were women, 3 were Asian, and 18 were White. All participants in this report were Cornell University undergraduates.

Apparatus. Stimuli were presented on Macintosh LC computers with 9-in. (22.86 cm) color monitors. Faces were presented with a resolution of 63 (horizontal [h]) × 90 (vertical [v]) pixels (approximately 23 × 32 mm) and were viewed from a distance of approximately 60 cm.

Stimuli. Stimuli were two realistic, gray scale, average male faces, one Black and one White. Each average was created from 16 faces by arbitrarily pairing them and producing composites of each pair using the Morph program (Gryphon Software). This program produces composites by distorting the shape of each face to an intermediate shape, then filling in average pixel intensities within the intermediate shape. The resulting eight composites were again paired to produce four composites. This process was repeated until a single composite of all 16 faces was completed. The faces in each average were taken from magazines and original photographs. Eight of the faces in each average were well-known individuals (actors and politicians), and 8 were not. For a more complete description of the stimuli and averaging process see Levin (1996). These averages were matched for mean brightness and contrast, so there was no difference between the races in terms of skin shading. An average of all 32 faces was created by combining the Black and White averages. The internal features (eyes, nose, and mouth) of this average were replaced with the internal features of the average for each race to create the final stimuli. Thus, each average had identical hair, ears, and jawlines and differed only in terms of internal features (see Figure 1).

Procedure. Participants were tested in groups of 4–7 on individual computers in the same room. The search task was run first.

For the search task, participants were instructed to detect the presence or absence of a given target face (the Black average or the White average) as quickly as possible without sacrificing accuracy. They were to hit the "1" key if the target face was present and the "2" key if it was absent. Instructions were given verbally; participants then read a similar set of instructions that included the Black and White target faces on the computers.

Search displays contained two, four, or eight faces presented against a white background. For the largest display size, faces were arranged in a diamond shaped pattern. (The total size of the largest displays was approximately 125 [h] × 110 [v] mm.) The target was equally likely to appear in each location. The same was true for smaller display sizes in which stimuli where arranged in symmetrical subsets of the locations in the largest display. The target face was present for half of the trials and absent for the other half. Distracter faces were multiple copies of the face that was not the current target. Thus, when the target was Black the distracters were White, and when the target was White the distracters were Black. On each trial, stimulus onset occurred within a single 16-ms refresh of the computer's monitor. Stimuli remained visible until participants responded. Feedback in the form of a "plus" for correct responses, and a "minus" for incorrect responses was presented immediately after every trial for 330 ms. Trials were separated by a blank-screen interval of 1,000 ms after the offset of the feedback stimulus.

After reading the first set of instructions, participants completed two blocks of trials, each containing 24 practice trials and 96 experimental trials. For each display size, there were 16 positive trials and 16 blank trials. After the first block of trials, the status of the target and distracter stimuli switched. If participants searched for the Black face among White distracters in the first set of trials, they searched for the White face among Black distracters in the second set. Task order was counterbalanced across participants.

After the search task, participants completed a perceptual identification task with the same two average faces. In this task, the faces were presented one at a time. In a simple go/no-go procedure, participants responded by hitting the "1" key as quickly as possible when the face on the screen was their target and did nothing when it was the other face. Again, there were

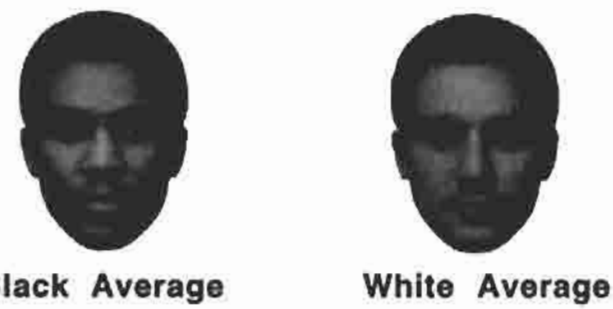


Figure 1. Average Black male and White male faces used in Experiment 1.

two sets of trials, one in which the Black face was the target and one in which the White face was the target. Alternate participants completed the tasks in different orders that were counterbalanced within each search task order. In each set of trials, the target face and the nontarget face were presented 15 times each.

The final task was the face recognition test. Participants were shown an inspection set of slides of 16 Black males and 16 White males from a high school yearbook. All pictures in the inspection set were masked so that most of the background and clothing were occluded. Slides were shown on a standard slide projector with an electronic advance and were visible for 3.2 s with a 0.8-s interstimulus interval (ISI). Participants were instructed to pay close attention because their memory for the faces would be tested. After viewing the inspection set, participants were instructed to record a "y" for each slide if they recognized it from the inspection set, and an "n" if they did not. Participants were also told to use their first impression and to avoid deliberating extensively about their response. The test set immediately followed these instructions and contained a total of 32 slides, 16 Black and 16 White. Within each group of 16, 8 were old slides from the inspection set, and 8 were new.

Results

Face recognition test. Face recognition errors were broken down by race of face, and type of error. Previous research (Levin, 1996; Ng & Lindsay, 1994; Valentine, 1991; Valentine & Endo, 1992) has shown that increased errors in CR recognition are often accounted for by increased false alarm rates. In the present case, the miss rates were similar between the races (Black: 61 total misses summed across participants; White: 73 misses), $F(1, 20) = 3.902$, $MSE = 0.879$, $p = .0622$, but there were more than twice as many false alarms for Black faces as for White faces (Black: 62 false alarms; White: 27 false alarms), $F(1, 20) = 10.115$, $MSE = 1.245$, $p = .0047$. Participants were therefore grouped on the basis of the difference in the number of false alarms between the races. Participants with the same number of false alarms to Black faces and White faces and participants who showed more false alarms to White faces were placed in the no-deficit group ($n = 9$); all other participants were placed in the deficit group ($n = 12$).

Search task. Data were entered into a 2 × 2 × 3 × 2 (Group [no-deficit/deficit] × Task [Black target/White target] × Display Size [2/4/8] × Target Presence [positive/blank]) mixed-factors analysis of variance (ANOVA) with group as the between-subjects factor.

A search asymmetry was observed that favored detection of Black faces. The Task × Display Size interaction was significant, $F(2, 38) = 6.873$, $MSE = 10{,}521$, $p = .0028$, and search slopes for Black targets (positive: 65 ms/item, blank, 170 ms/item) were less steep than slopes for White targets (positive: 88 ms/item, blank: 175 ms/item). This interaction was further modified by a significant Task × Display Size × Group interaction, $F(2, 38) = 3.619$, $MSE = 10{,}521$, $p = .0364$ (see Figure 2). Separate Task × Display Size × Target Presence ANOVAs on each group revealed a small and marginally significant search asymmetry for the no-deficit group: Task × Display Size interaction, $F(2, 16) = 3.490$, $MSE = 5{,}839$, $p = .0552$. Those in the deficit group, however, showed a significant asymmetry favoring detection of Black faces:

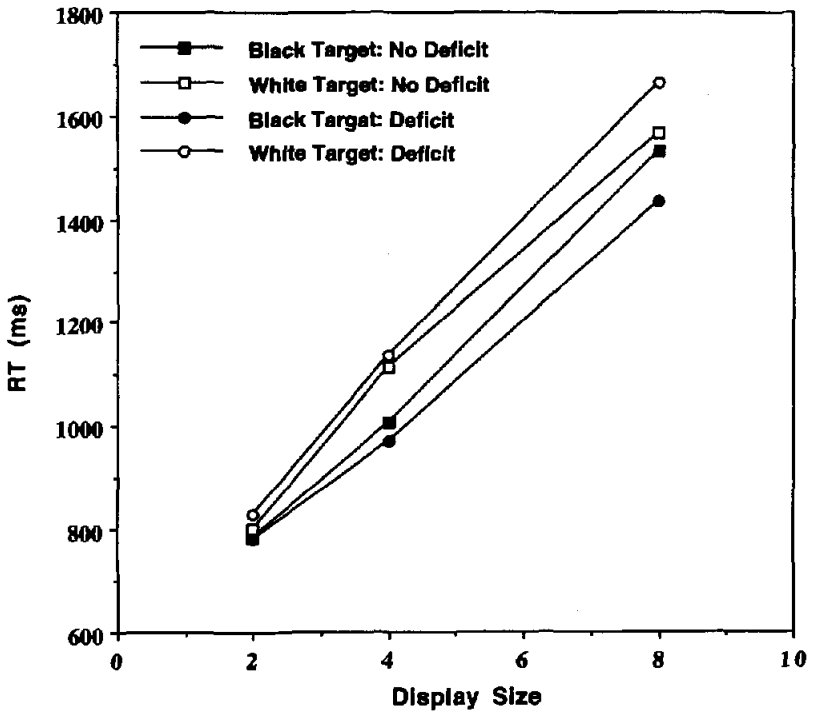


Figure 2. Mean search reaction times (RTs) for Black and White target searches in participants who do and do not show a cross-race recognition deficit.

Task × Display Size interaction, $F(2, 22) = 7.296$, $MSE = 13{,}927$, $p = .0025$.[2]

Perceptual ID task. Data from the perceptual ID task were entered into a 2 × 2 (Group [no deficit/deficit] × Race of Face [Black/White]) mixed-factors ANOVA with group as the between-subjects factor. Neither main effect approached significance, but the Group × Task interaction was nearly significant, $F(1, 19) = 3.852$, $MSE = 911.2$, $p = .0645$. Simple effects analysis shows that Black faces were detected significantly more quickly than White faces by the deficit group (Black target: 443 ms; White target: 472 ms), $F(1, 19) = 5.493$, $MSE = 911.2$, $p = .030$, whereas there was no difference in detection times for the no-deficit group (Black target: 452 ms; White target: 444 ms, *ns*).

Discussion

Experiment 1 replicated two critical aspects of Levin's (1996) Experiments 5 and 6. First, a search asymmetry that favored detection of Black faces was again observed. Second, participants who showed a CR recognition deficit identified the Black face more quickly in the perceptual identification task. This helps to rule out explanations for the asymmetry that posit a serial search in which participants' expertise at identifying individual White faces allows them to quickly identify and reject the White distracters. Further, the perceptual identification advantage for Black faces was not observed for participants who showed no CR recognition deficit.

These findings also support the proposed relationship between the feature-selection asymmetry and the CR recognition deficit. Participants who do not show the typical CR recognition deficit show a smaller search asymmetry. Although this evidence is only correlational, the same could be said for a number of well accepted explanations for the deficit that rely upon correlations between recognition performance and some aspect of participants lives. Added to Levin's (1996) finding that the asymmetry is smaller among African Americans and nonexistent among African Nationals, these data give the feature-selection explanation for the CR recognition deficit a viability that at least matches the contact hypothesis.

In addition to supporting the feature-selection hypothesis, the search-recognition relationship helps eliminate the concern that the search asymmetry is caused by low-level perceptual difference between Black faces and White faces. If Black faces (or the particular average used here) were easier to find because some basic perceptual feature makes them stand out, then all participants should show the asymmetry, regardless of how accurately they recognize Black faces.

However, a number of explanations for the CR recognition deficit also specify coding differences between CR and SR faces and therefore go beyond a correlational analysis of recognition skills. In particular, some expertise-based accounts of the recognition deficit propose that differences in coding the configuration of facial features account for the recognition deficit. Experiment 2 was performed to test whether this kind of coding difference could also explain search asymmetry.

Experiment 2

One of the major differences between CR and SR face coding is that people appear less likely to use configural information to code CR faces. Two findings support this hypothesis. First, the *inversion effect,* whereby face recognition is more disrupted by inversion than is recognition of other objects (Yin, 1969), may be smaller for CR faces. Rhodes, Tan, Brake, and Taylor (1989) tested White participants' recognition of upright and inverted White and Asian faces. They found that inversion disrupted recognition of Asian faces less than White faces (but see Valentine & Bruce, 1986). Given that inversion is thought to differentially disrupt expertise-based configural coding (Diamond & Carey, 1986; Farah, Tanaka, & Drain, 1995; Martini & Nakayama, 1998; Rhodes, Brake, & Atkinson, 1993), this is evidence that configural coding was less prominent for Asian faces. Fallshore and Schooler (1995) provide converging evidence for reduced configural coding in CR faces. Their study was based on the *verbal overshadowing effect,* in which verbalizing a facial features reduces recognition accuracy as compared with no rehearsal at all. Schooler and Engstler-Schooler (1990) argued that verbal overshadowing occurs because verbalizing faces causes inaccurate part-based information to overwrite more accurate configural information. Fallshore and Schooler reasoned that if CR faces are not coded configurally then they should show no verbal overshadowing effect. As expected, they found that verbalization did not affect CR recognition accuracy, again suggesting that configural processing is absent in CR faces.

If configural processing causes differential recognition performance, then it may underlie differential search performance as well. There is, however, reason to believe that this is not the case. Levin (1996) used a simple classification task and found that CR faces are classified by race more quickly than SR faces, and further that this difference is also observed for inverted faces. This finding came as a surprise because it suggests that there is a substantive difference in the way we process CR and SR faces that is not caused by configural coding. If the CR classification advantage is based on the same feature-selection process as the search asymmetry, then the inversion finding implies that the search asymmetry does not depend on configural coding. This is, however, a bit premature because inversion has not been attempted with the visual search task.

In Experiment 2, therefore, participants searched for inverted faces. If the search asymmetry is not due to differential configural coding then it should be observed in inverted stimuli as well. If differential configural coding between the races does cause the asymmetry, then inversion should effectively eliminate it.

Method

Participants. A total of 18 participants completed Experiment 2. Of these, 4 who did not meet the 5% error criterion were dropped from the analysis. The average error rate of those remaining was 2%. Of the remaining 14 participants, 6 were men, 3 were Asian, and 11 were White.

Stimuli and procedure. The procedure was similar to that in Experiment 1 with the following exceptions. First, display set sizes were reduced,

[2] Because small nonlinearities appear to drive the nearly significant Task × Display size interaction for the no difference group, the ANOVA for both groups was rerun with the search slopes as a dependent variable. The results of this analysis are essentially the same as before. Both the task effect and the Group × Task interaction were significant (*ps* = .0360, .0452, respectively).

and stimuli were enlarged. Participants searched displays containing two, three, or six stimuli. The same gray-scale faces as in Experiment 1 were presented with a resolution of 80 (h) × 110 (v) pixels (approximately 29 [h] × 40 [v] mm). The average faces were arranged in a symmetrical rectangular formation with two rows of up to three faces each. Displays were approximately 124 (h) × 98 (v) mm. These changes were deemed necessary because pilot tests using inverted stimuli identical to those in Experiment 1 produced high error rates. Another change was that the number of trials per cell was increased from 16 to 24, for a total of 288 trials. (This was done simply to increase experimental power once it became clear that participants had no problem attending for this long.) Test trials for each task (Black target/White target) were preceded by 18 practice trials. Finally, the perceptual ID task was dropped from the protocol. Faces were inverted by rotating them 180°.

Results

Data were entered into a 2 × 3 × 2 (Task [Black target/White target] × Display Size [2/3/6] × Target Presence [positive/blank]) within-subject ANOVA. The Task × Display Size interaction was significant, $F(2, 26) = 6.940$, $MSE = 11{,}932$, $p = .0038$. As can be seen in Figure 3, visual search for the Black target was faster (positive slope: 62 ms/item; blank slope: 153 ms/item) than for the White target (positive slope: 93 ms/item; blank slope: 195 ms/item); task effect on slopes, $F(1, 13) = 7.636$, $MSE = 2{,}470$, $p = .0161$.

Discussion

Even with inverted faces, the search asymmetry favoring detection of Black faces remained. This finding suggests that differential configural coding between SR and CR faces is not responsible for the search asymmetry. If it were, then the inversion manipulation would be expected to erase this difference and therefore eliminate the search asymmetry. What, then, does this imply for the information that drives the search asymmetry? It suggests that race-specifying information is processed in a manner analogous to a facial part or feature and is not coded as a relationship among features. Although the positive finding in inverted faces might be interpreted to mean that race-feature coding is not part of normal face processing and is perhaps based on a low-level feature, recent findings suggest that at least some kinds of face-specific information are codable in inverted faces. For example, Bartlett and Searcy (1993) found that facial expression coding is unaffected by inversion and further that isolated parts are also coded reasonably accurately in inverted faces. This suggests that inversion is quite specific in its disruption of configural coding and, therefore, that the asymmetry could easily be based on face-specific information.

Given that the search asymmetry remains for inverted faces, its status as a substantive difference between CR and SR face coding becomes separated from other findings differentiating these processes. Therefore, the idea that the CR recognition deficit is caused by an inability to adequately code facial configuration might now have a competitor. Experiments 3 and 4 further call into question encoding-expertise explanations for the deficit by using a perceptual discrimination task to show that participants are better able to discriminate among CR faces than among SR faces when the task selects race-specifying information.

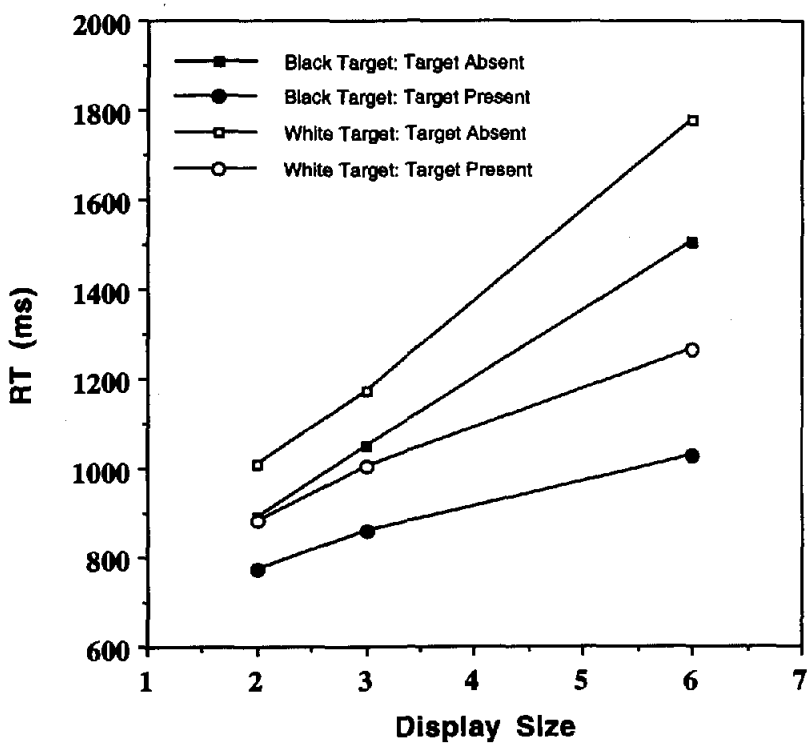


Figure 3. Mean search reaction times (RTs) for inverted Black and White targets.

Experiment 3

In Experiment 3, one of the more counterintuitive predictions of the feature-selection hypothesis was tested. That is, perceptual sensitivity should be enhanced among Black faces as compared with White faces when the discriminations depend on race-specifying information. If the deficit is associated with a bias to code race-specifying information in CR faces, then discrimination should be relatively accurate for stimuli that differ in the degree to which they instantiate CR race-specifying information. In this experiment, participants discriminated between subtly different faces distorted to vary on a Black-to-White shape continuum (see Figure 4). Thus, faces at the Black end of the continuum range from the average Black face to faces that have less information specifying "Blackness" because they look more like the White face, whereas faces in the White end of the continuum vary in the same way, and to the same degree. If participants recognize Black faces poorly because they emphasize race-specifying information, then they should be paradoxically accurate when making discriminations in the Black end of this continuum. This prediction conflicts with most encoding expertise hypotheses, which assume that people cannot make subtle perceptual discriminations among CR faces.

Experiment 3 tests for improved race-based discrimination among Black faces by borrowing a paradigm from research on categorical perception in faces. In these experiments (i.e., Beale & Keil, 1995; Etcoff & Magee, 1992; Levin & Beale, 2000), participants viewed pairs of stimuli at different locations along a continuum between two end points. In Beale and Keil, for example, the end points were familiar faces, and the continuum was made by using the previously described morphing algorithm to create controlled blends of the end points. Participants attempted to discriminate which of a pair of faces from the continuum was more like one of the end points. For example, on a given trial, they might view a face representing a 60% John F. Kennedy/40% Bill Clinton blend next to a 40% Kennedy/60% Clinton blend and be asked which looks more like Kennedy. In the second task, participants classified each intermediate for its perceived likeness to one end point or the other. This produces a classification boundary between faces that are usually classified with the first end point and faces usually classified with the second. Although previous research on categorical perception tested for more accurate discrimination at category boundaries, here the focus is on the converse question: How accurate are participants at making discriminations within each of the continuum end regions (e.g. faces that are consistently classified as Black vs. faces that are consistently classified as White)?

Method

Participants. Thirteen participants completed the present experiment. Of these, 7 were White, 3 were Hispanic, 3 were Asian, and 4 were women.

Stimuli. A stimulus continuum between the Black average face and the White average face was created with the distortion algorithm described earlier. This process involves creating intermediate shapes based on key-point maps that mark the correspondence between relevant facial features in each end-point face. Once the intermediate shape is calculated, luminance values of the end points are blended in proportion to the proximity of the intermediate to each end point. For example, a 20% Black-White intermediate is created by distorting the shape of the Black face to be 20% of the distance from the Black face to the White face. Luminance values for the intermediate would be a blend including 80% of the Black face, and 20% of the White face. In this way, sets of stimuli including continuum end points and intermediates representing 10% steps between the end points were created (Figure 4 includes a sample from this continuum). For convenience, the continuum is described as starting at the Black end point (0%) and ending at the White end point (100%). These stimuli, like the ones used previously, were matched for average luminance and contrast.

Procedure. Participants completed a block of discrimination trials followed by a block of categorization trials. At the start of the discrimination trials, they viewed the continuum end points (undistorted Black and White averages) and were instructed that they would be making discriminations among faces similar to these. Trials consisted of two events: First, participants were shown an instruction asking them to indicate which of the upcoming two faces was more like one of the end point faces. On half of the trials, they were asked which face looked more like the Black face, and on the other half they were asked which looked more like the White face. After pressing the space bar to indicate their readiness to continue, participants viewed two faces separated by 20% of the total continuum. Faces were presented side by side with a resolution of 220 (h) × 315 (v) pixels (approximately 80 [h] × 112 [v] mm) for 1 s. After the stimuli were erased from the screen, participants responded by typing a "1" on the computer keyboard if they thought the left face was more like the designated end point and a "2" if they thought the right face was more like the designated end point. Each of the nine discrimination pairs were presented a total of eight times, counterbalanced for screen side and target response instruction.

A categorization task followed the discrimination task so that the boundary between different regions on the continuum could be objectively determined. First, participants again viewed the two continuum end points. They were instructed that they would see a series of faces similar to these and to indicate which end point each face most resembled. Faces were presented one at a time in response-terminated displays. The 11 stimuli (the 9 blends and the 2 end points) were each presented eight times for a total of 88 randomly ordered trials.

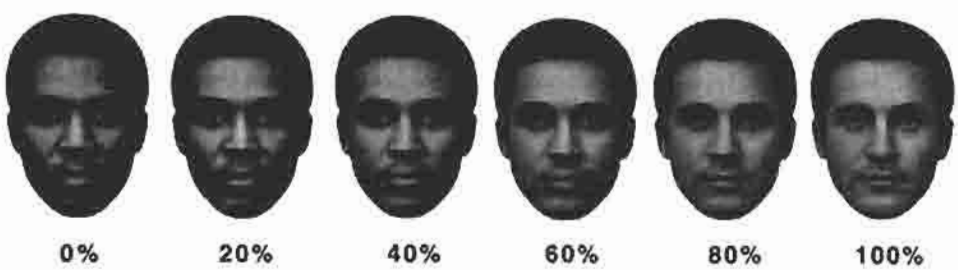


Figure 4. Black–White continuum. Only 20% increments are shown here.

Results

Categorization task. Results from the categorization task were used to isolate within-category discrimination pairs. The selection criterion was essentially the reverse of that used by Etcoff and Magee (1992) and Beale and Keil (1995) to select cross-boundary pairs (i.e., pairs for which each face was perceived as most similar to different end point faces). Pairs for which both faces were classified as similar to the same end point on more than 66% of trials were selected to represent each face category. This scheme selected the 0–20%, 10–30%, and 20–40% pairs for the Black category, and the 60–80%, 70–90% and 80–100% pairs for the White category.

Discrimination task. On the basis of the above selection, discrimination accuracy within each continuum end region was calculated for each participant by averaging discrimination scores across all within-category pairs. These data were entered into a within-subject ANOVA with continuum end (Black/White) as the single factor. Discrimination was more accurate at the Black end of the continuum (83%) than at the White end (69%), $F(1, 12) = 12.267$, $MSE = .012$, $p = .0044$).

Discussion

Experiment 3 shows that it is possible to observe more accurate perceptual discrimination among CR faces than SR faces. This finding is compatible with the feature-selection hypothesis, under the assumption that variation in a selected feature allows improved feature-relevant discrimination, even when that variation occurs in CR faces.

There is, however, another plausible explanation for the findings of Experiment 3. It is based on the assumption that accurate Black end region discrimination is not a question of increased sensitivity to variation in CR race-specifying information but is instead a more global effect in which discrimination among variants of Black faces is more accurate than discrimination among variants of White faces. This alternative is based on the perceptual magnet hypothesis whereby "strong" representations assimilate near variants (Iverson & Kuhl, 1995; Kuhl, 1991). In Iverson and Kuhl's formulation, a strong representation causes reduced sensitivity to variation within the same perceptual category. Applied to the present data, this means that a strong White face representation causes reduced sensitivity to distortions of that face.

Experiment 4

According to the perceptual magnet hypothesis, discrimination is less accurate for stimuli that are similar to strong representations. Therefore, the discrimination advantage for Black faces observed in Experiment 3 could be due to globally reduced accuracy in discriminating variants of White faces, assuming the White faces to be associated with stronger representations. Experiment 4 was therefore necessary to eliminate this alternative explanation. This was done by comparing accuracy of discrimination for Black and White faces along continua between two faces of the same race (i.e., Black–Black and White–White continua), with accuracy along between-race continua (i.e., Black–White). If the feature-selection hypothesis is correct, sensitization should occur only for race-based discriminations between face pairs that vary in terms of Black race-specifying information and should not occur for Black end regions on the Black–Black within-race continua. The perceptual magnet hypothesis predicts that discrimination should be more accurate in general near Black end points whether or not they occur along a between-race continuum. This design also reduces the concern that stimulus-specific factors are responsible for the observed sensitization because it replaces the average faces with different individual faces along a total of 12 different between-race and within-race continua.

This experiment also dispensed with instructions that directly invoke a representation of the continuum end points. Instead of indicating which of two faces look more like one of the end points, a more traditional ABX task (e.g., Liberman, Harris, Kinney, & Lane, 1961) was employed. In this task, participants first saw the discrimination pair (A and B; both images at the same time), followed by the criterion (X) image. Participants then responded by indicating which of the first two images (A or B) matched the third image (X). After completing the discrimination task, participants completed a set of three brief categorization tasks similar to those in Experiment 3.

Finally, Experiment 4 included the same face recognition test used in Experiment 1. This allowed a test of the hypothesis that improved discrimination accuracy in the Black end region of between-race continua should only be observed for those participants who show a CR recognition deficit.

Method

Participants. A total of 20 participants (11 men) completed Experiment 4. Of these, 17 were White, 1 was Asian, 1 was Hispanic, and one gave an uncodable response.

Stimuli. Sixteen unfamiliar male faces (8 Black and 8 White) were used as stimuli in this experiment. They are equated for average distinctiveness (and rated familiarity) across the races (see Levin, 1996). The 8 faces from each race were paired among themselves to make 4 continua within each race. In addition, 4 of the faces in each race were paired between race categories for a total of 12 continua. To equalize similarity between continuum end points in each of the three sets of continua (Black–Black, White–White, and Black–White), 10 judges rated the similarity of all possible pairings among the 16 faces. Judges were instructed to rate similarity on the basis of overall similarity of face shape by means of a 7-point scale (1 = *very similar*, 7 = *very dissimilar*) and were further instructed to ignore differences in race when viewing CR pairs. On the basis of these ratings, sets of face pairs were chosen to equalize average similarity ratings between the three stimulus groups (mean similarity ratings: Black–Black pairs, 4.15; White–White pairs, 4.08; Black–White pairs, 4.05; *SD* for all chosen pairs = 0.5823). Once the set of face pairs was selected, the method described in Experiment 1 was used to create continua with 10% increments between the members of each pair. The 12 continua were arbitrarily grouped into four stimulus sets, each containing one between-race pair and two within-race pairs (one Black–Black, and one White–White). Each participant saw one of these four stimulus sets. Thus, the end point stimuli for the between-race continuum for a given participant served as end points on within-race continua for other participants. Stimuli were presented at the same size and resolution as in Experiment 3.

Procedure. Participants completed a set of discrimination trials and three sets of classification trials, one for each face pair. In the discrimination task, participants first saw two faces for 1,000 ms; then, after a 500-ms delay, they saw a third face that matched one of the first two. Participants indicated which of the first two faces matched the third face by pressing the "1" key for the left one, and the "2" key for the right). In all cases, the two

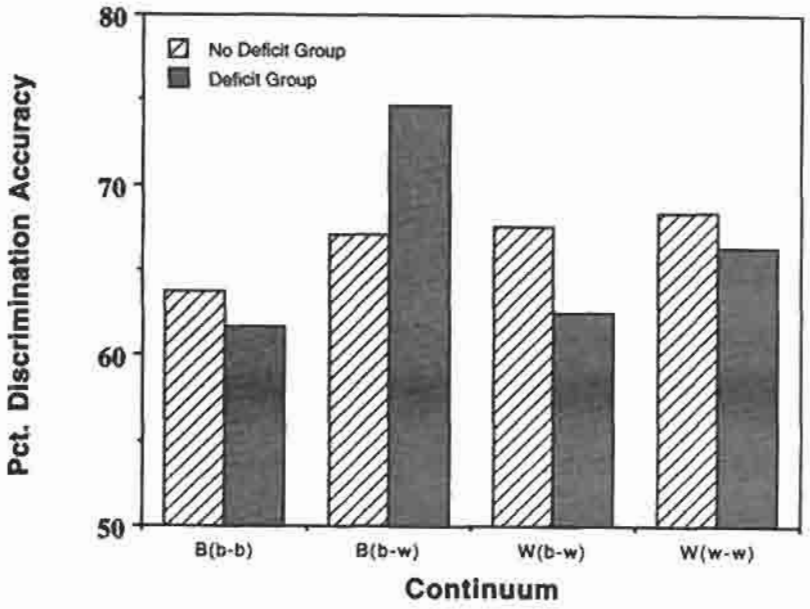


Figure 5. Mean percentage (Pct.) of discrimination accuracy by race and continuum. End points are denoted by the first capital letter, and the continua are denoted by the lowercase letters (b-b and w-w are within-race continua for Black and White faces, respectively; b-w are between-race continua.)

faces in the initial display were separated by two steps, or 20% of the total continuum. The discrimination trials for each of the three continua were randomly mixed in a single block of trials. There were 8 trials (counterbalanced for screen side and correct response) for each of the nine face pairs along three continua, for a total of 216 trials.

After completing the discrimination trials, participants completed three blocks of classification trials in which they classified individual faces along one of the continua from the discrimination trials. At the beginning of each block, the two end point faces were presented and paired with an arbitrarily chosen name. For each trial, participants instructed to hit the "1" key if they thought the face was more similar to one of the end point faces and to hit the "2" key if it was more similar to the other. Each of the 11 faces from the continuum was presented three times, for a total of 33 trials in each of the three blocks. The order of blocks was counterbalanced across participants such that half saw the between-race pair first, and half saw the within-race pairs first (in two successive blocks). The order of the two within-race pairs was also counterbalanced. Finally, participants completed a recognition task identical to that in Experiment 1.

Results

Recognition task. The results of the recognition task were almost identical to those of Experiment 1. Participants made nearly twice as many false alarms to Black faces (38) than to White faces (20), $F(1, 19) = 7.364$, $MSE = 1.100$, $p = .0138$, whereas the number of misses was similar between the races (Black: 66; White: 55; $F < 1$). As in Experiment 1, participants who made the same number of false alarms to Black and White faces or who made more false alarms to White faces were placed in the no-deficit group ($n = 10$); the rest were placed in the deficit group ($n = 10$).

Classification task. The 40–60% pair was selected as the cross-boundary pair in all three conditions, according to the same criteria as in Experiment 1. Again, this caused the 0–20%, 10–30%, 20–40%, 60–80%, 70–90%, and 80–100% discrimination pairs to be selected as end region trials.

Discrimination task. Mean discrimination accuracy was computed for end region stimuli along Black–Black, White–White, and Black–White continua.[3] Comparisons between end regions along within-race and between-race continua were computed only for the four stimuli of each race that appeared in both conditions. This allowed a comparison of discrimination accuracy for end regions near the same faces, but along different continua.

Average discrimination accuracy for between-race continua were entered into a 2 × 2 (Continuum End [Black/White] × Recognition Group [deficit/no deficit]) mixed-factors ANOVA with recognition group as the between-subjects factor. This produced a significant Continuum End × Recognition Group interaction, $F(1, 18) = 5.336$, $MSE = 0.007$, $p = .0330$; participants who showed a CR recognition deficit also showed significantly improved discrimination at the Black end of Black–White continua (75% for the Black end vs. 63% for the White end), $F(1, 18) = 9.972$, $MSE = 0.007$, $p = .005$ (see Figure 5), whereas participants who showed no CR recognition deficit showed equally accurate discrimination at both ends of the continua (68% for the Black end vs. 67% for the White end; *ns*).

Data from all four continua for the deficit group were entered into a single factor end region (Black end, within-race/Black end, between-race/White end, between-race/White end, within-race) ANOVA to test the perceptual magnet hypothesis. The end region

[3] For the sake of completeness, categorical effects were also evaluated by contrasting discrimination accuracy for the category boundary pair with accuracy for the other eight pairs. These contrasts were nonsignificant for within-race continua: Black–Black, $F(1, 19) = 1.240$, $MSE = 0.038$, $p = .2793$; White–White, $F < 1$. The contrast for the between-race continuum, however, was nearly significant, $F(1, 19) = 4.145$, $MSE = 0.025$, $P = .0559$.

effect was significant, $F(3, 27) = 3.082$, $MSE = 0.011$, $p = .0441$, and pairwise comparisons indicate that discrimination was significantly more accurate for Black end regions on Black–White continua (75%) than for end regions near the same Black faces when they were part of Black–Black continua (62%; $p < .05$, Duncan test). Thus, the pattern of sensitization was limited to the Black end of between-race continua.

In addition, average discrimination accuracy was calculated for all pairs of Black versus White faces on within-race continua. These were entered into a two-factor (Race [Black/White] × Recognition Group [deficit/no deficit]) mixed-factors ANOVA with recognition group as the between-subjects factor. Discrimination was more accurate overall for White within-race continua (71%) than for Black within-race continua (67%), $F(1, 37) = 5.137$, $MSE = 0.004$, $p = .0360$. This difference in discrimination accuracy did not, however, interact with the presence of the CR recognition deficit, and the main effect for recognition group was not significant (Fs < 1).

Discussion

The results of Experiment 4 support the feature-selection hypothesis by showing a pattern of sensitization that limits itself to the Black end region on between-race continua. This is counter to the perceptual magnet hypothesis, which makes no theoretical distinction between within-race and between-race continua. It only specifies that stimuli similar to a well known exemplar should be difficult to discriminate. Further, if the perceptual magnet effect were responsible for suppressing discrimination accuracy near White faces, then discrimination accuracy should be reduced for White within-race end regions as compared with Black within-race end regions. This was not the case. In fact, the data show a trend in the reverse direction. Not only are these data counter to a perceptual magnet explanation of Black end region sensitivity, but they also run counter to any explanation of the CR recognition deficit that depends on globally reduced perceptual sensitivity for CR faces. This will be explored more in the general discussion.

In addition, Experiment 4 revealed a relationship between the presence of the CR recognition deficit and improved discrimination among faces near the Black end region of the Black–White between-race continuum. Therefore, individuals who show the CR recognition deficit may do so because they tend to select race-specifying information when coding CR faces. This kind of sensitivity leads to a paradoxical reversal of the recognition deficit—better discrimination accuracy among CR faces as compared with SR faces so long as the discrimination task selects race-specifying information. The accuracy levels for the within-race continua also suggest that, when forced to do so by task constraints, individuals who show the deficit are able to discriminate between CR faces as accurately as those who show no deficit. This supports the hypothesis that the recognition deficit, at least as measured here, may not reflect a perceptual encoding limit but is instead the result of a feature-selection process whereby some individuals are oriented toward coding information relevant to race.

Before going on to the general discussion, it is important to address two final methodological issues. First, in all of the present experiments and in Levin (1996), increased detectability and discriminability of CR faces could be considered a relative effect that may overlay some basic perceptual feature that makes Black faces easier to detect. In the experiments reported here, for example, White participants generally were more able to detect Black faces, and discriminate variants of "Blackness." Within this effect was an interaction between ability to recognize Black faces and ability to discriminate among them or detect them. This interaction is the central element in this article; it follows from the prediction that the CR face category is more salient than the SR face category for majority group subjects. However, it remains possible that some perceptual feature specific to Black faces also makes them more detectable or discriminable. If this were true, the basic feature-selection hypothesis would remain intact because it depends on the interaction effect, but it would take a slightly different form. Consider the discrimination tasks in which no-deficit White participants discriminated between Black and White variants with equal accuracy. If one were to argue that a stimulus artifact drives the basic improvement in discrimination accuracy for Black faces, then one would have to assume that White participants who accurately recognize Black faces do so by suppressing this feature. Thus, the feature selection would still be the root of the CR deficit, but it would be a selection in which no-deficit participants deselect race-specifying features, whereas deficit participants default to salient perceptual features. The findings would still conflict with the encoding expertise explanation of the deficit because improved recognition is still not associated with globally improved encoding accuracy—it is associated with lower accuracy for the category-spanning Black end regions.

Second, the issue of generalizability deserves some comment. In all experiments reported here, White participants' recognition of Black faces has been the sole example of CR recognition. Although generalization to other populations would be beneficial, the essential point of this paper is that the dominant explanation of the CR recognition deficit has difficulty accommodating any situation in which discrimination between CR faces is accurate. Thus, these experiments relied on the most frequently tested SR–CR contrast to reveal an exception to the prevailing theory and suggest that we need to reconsider expertise hypotheses, whether or not the findings generalize. That said, it does appear that similar processes occur for other groups of faces. For example, Valentine and Endo (1992) found that British participants classified Asian faces faster than White faces. If this CR classification advantage is caused by similar processes as the CR classification advantage reported by Levin (1996) for Black faces and the CR search and discrimination advantages reported here, then it is likely that these effects will generalize to other SR–CR contrasts.

General Discussion

These experiments have explored a natural object classification: determining the race of a face. In doing so, they suggest that the encoding expertise explanation of the CR recognition deficit is incorrect by showing that participants who are poor at recognizing Black faces are paradoxically quick to detect them and accurate in discriminating among them. First, in Experiment 1 the visual search asymmetry favoring CR faces was replicated and related to CR face recognition accuracy. Participants who showed no CR recognition deficit also showed no search asymmetry. In Experiment 2, the asymmetry remained when inverted faces were used as stimuli. This suggests that configural coding differences between CR and SR faces do not account for the asymmetry. In Experi-

ments 3 and 4, discrimination was more accurate for faces that were similar to Black end points, but only when these faces fell on a between-race continuum, and only in individuals who showed the CR recognition deficit.

The general discussion proceeds by first examining the specific relationship between these findings and existing explanations of the CR recognition deficit. The second section discusses the implications of these results more generally for models of object classification. It suggests that a true understanding of object categories should include not only an understanding of perceptual encoding and abstract similarity relationships among representations, but also an explicit account of the specific cognitions that constrain these perceptual processes in any given domain.

Implications for Explanations of the CR Recognition Deficit

The present findings have a number of interesting implications for explanations of the relationship between face categories and face recognition. In particular, these findings appear to conflict with several expertise-based explanations of the CR recognition deficit. In one variant, differences in encoding expertise are embodied by the assumption that CR faces are globally more psychologically similar to each other than SR faces and therefore are harder to discriminate (Valentine, 1991; Valentine & Endo, 1992). Another hypothesis is based on neural net simulations. In this case, a net can be trained on a series of faces dominated by one race or another. When the net must reconstruct previously seen faces, it is less accurate for faces from the infrequently encountered group (O'Toole, Deffenbacher, Abdi, & Bartlett, 1991). Most other explanations of the CR recognition deficit assume the same thing, although they are usually not as specific as the above two examples. What these explanations share is the assumption that broadly applicable principles of perceptual encoding and similarity among face exemplars (i.e., individual faces represented in memory) serve as a substantive way of understanding the CR deficit. I argue that this kind of model is inherently ill equipped to handle the data presented here.

At the simplest level, similarity and encoding expertise hypotheses plainly conflict with the finding that discrimination accuracy is, in fact, greater among variants of CR faces when race-specific information is selected by the task. To get a sense of the conflict between this finding and encoding-expertise explanations of the CR deficit, it is useful to consider more closely Valentine's (1991) multidimensional space model. Intended as a general framework for understanding face recognition, the multidimensional space model assumes that faces are recognized and classified by referring percepts to a representational space in which exemplar computations determine the ease of recognizing and classifying faces. Under this scheme, recognition is equated with matching the current percept to a node representing a previously seen face. This process is facilitated if the percept maximally activates one node to the exclusion of others. Where activations are less specific, recognition is more difficult. Therefore, regions of face space that are dense with many exemplars per unit area are characterized by inaccurate recognition. Face classification is explained by assuming that classification is fast when a precept activates a large number of exemplars from a given category. If one assumes that CR faces are represented in a "dense" region of this representation (see top of Figure 6), then classifying CR faces by race should be faster than classifying SR faces because of the same increased density that putatively makes CR recognition difficult, a prediction confirmed by Valentine and Endo (1992) and replicated by Levin (1996). Thus, the multidimensional space model assumes that CR faces are more psychologically similar to one another than SR faces, and therefore should be less discriminable and more classifiable overall.

Although the discrimination prediction conflicts with the present data, consider the most plausible fix to such a model. One could simply assume that although the CR face space is generally more dense than the SR face space, it is elongated along the CR–SR axis, as shown in Figure 6. Then, with the addition of a selective attention mechanism, one could assume that participants attend to race-specifying information (as represented on the elongated CR–SR axis) when the A and B stimuli (in the ABX task) differ on the elongated race axis. This would allow for improved discrimination among CR faces when race-specifying information is selected while retaining the globally increased density of the CR face space when putatively orthogonal individuating information is attended to. However, this fix conflicts with the idea that CR faces are more quickly classified than SR faces. If the CR–SR axis is elongated, then the feature selections in a race classification task should also occur in this elongated and less dense space. Under the multidimensional space model, this should lead to slower CR classification, not faster classification as has been repeatedly observed (Levin, 1996; Valentine & Endo, 1992).

The model could be modified yet again to assume that the face space is mutable across tasks. Perhaps CR face space is elongated for discrimination tasks and spherical for classification tasks. This

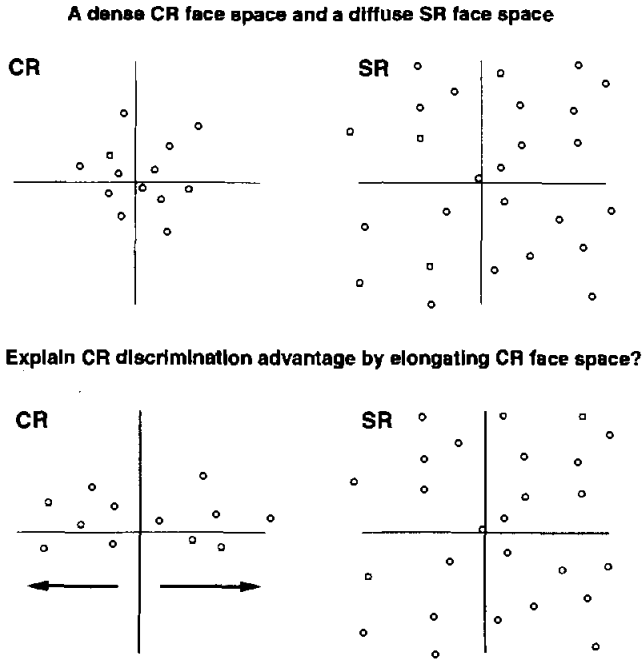


Figure 6. Fixing a multidimensional scaling model to account for facilitated discrimination in Black face end regions. CR = cross-race; SR = same race.

modification is, however, unsatisfying. First, the model is playing catch-up to empirical findings rather than making predictions. Second, and perhaps more problematic, is that this kind of mutability vacates the predictions that the model made in its original form. Given that the multidimensional space model is intended to allow simple abstract rules to predict performance across situations and tasks, the possibility that face space is forced to change across just those tasks is deeply problematic.

Although a simple perceptual encoding expertise model appears inadequate to explain the CR deficit, perhaps a more sophisticated one might. One possibility would be to explicitly include feature-selection differences between expert SR and inexpert CR face coding as aspects of encoding expertise. This approach has been used previously to account for visual expertise by Myles-Worsley, Johnston, and Simons (1988) who found that expert radiologists were less accurate than nonexperts at remembering nonpathological variation in chest x-rays. They argue that the experts attend more closely to pathological variation than nonexperts but that they attend less to other information than nonexperts. In this case, a large part of expertise is knowing how to efficiently distribute attention to the correct task-relevant features. Although the feature-selection difference in radiologists depends on many years of explicit practice, a factor that distinguishes it from the feature-selection difference in faces (which can be quickly erased), other findings suggest that visual expertise can be associated with quickly learned feature selections. For example, Biederman and Shiffrar (1987) studied chick-sexing experts who must discriminate subtle differences in the genitals of male and female day-old chicks. They found that college students could be trained to sex chickens with accuracy comparable to that of experts once the key features were pointed out to them. As such, the feature-selection hypothesis is not completely inconsistent with current accounts of visual expertise. Instead, it runs counter to the more specific hypothesis that the expertise involves an ability to make quantitatively more fine perceptual discriminations among SR faces, as reflected in metric differences in CR and SR face spaces.

Accordingly, the feature-selection hypothesis can coexist with other known CR–SR face processing differences. In particular, findings indicating less configural coding in CR faces (e.g., Fallshore & Schooler, 1995; Rhodes et al., 1989) would be explained in terms of feature selection. Participants do not code the configuration of CR face features but code nonconfigural race-specifying features instead. However, the default version of this hypothesis is that participants fail to code configuration because they can't, whereas the feature-selection version implies that they simply don't. Therefore, the feature-selection hypothesis would predict that participants will code the configuration of CR faces given the correct task. It is interesting to note that existing data do hint that this is the case. Although Rhodes et al. found a smaller inversion effect for CR faces (indicating less configural coding because inversion specifically disrupts configural coding), Valentine and Bruce (1986) found a greater inversion effect for CR faces than for SR faces. The difference between the two studies is that Valentine and Bruce allowed participants 5 s to view each CR inspection set face and only 2 to view the SR faces, whereas Rhodes et al. displayed the faces for the same amount of time. Therefore, if participants have sufficient time, they may go beyond the default feature selection and include configural features even in CR faces. Clearly, for the feature-selection hypothesis to be more strongly supported, we should expect to see equivalent configural coding in CR and SR faces even with matched exposures. However, the more general point remains: A lack of perceptual expertise does not absolutely prevent configural coding in CR faces. Therefore, the feature selection itself, and the social cognitions that maintain it, may be more central in explaining the deficit than limiting differences in perceptual expertise.

The Role of Similarity in Object Recognition

If the above analysis is correct, then similarity-based exemplar models of face categories are insufficient to account for basic face-processing effects. Without some means of specifying what features enter into the similarity comparison, it is difficult to predict performance across task contexts and populations. This problem is, however, not limited to faces. If the feature-selection process described in faces can occur in other domains—and there is no reason to assume it cannot—then any model of object recognition and classification that ignores this selection process may be incomplete. Such models are common and usually are instantiated in terms of low dimensional representations of perceptual space (see for example, Edelman, 1995; Medin & Schaffer, 1978; Nosofsky, 1984). This section borrows from the concept literature to discuss possible solutions to the problems inherent in using similarity to understand object categories. Some of these solutions seek to constrain similarity-based models with a systematic understanding of task contexts; others emphasize the need to understand more abstract cognitions associated with specific categories. Although much of this research is based on deliberated decisions about category membership, similarity also needs to be constrained for fast, undeliberated object recognition and classification.

Research in the past 10–15 years has led most investigators to conclude that similarity-based models of concepts (in the form of simple exemplar or prototype models) fail to explain many important aspects of human concept usage (Medin, 1989). One of the most important findings from this literature is that participants' classification and similarity judgments seem to confound simple metrics of physical similarity. For example, Rips (1989) demonstrates that participants consider it more likely that a 3-in. disk is a pizza than a quarter despite the fact that its most salient feature, size, makes it more similar to the quarter. Thus, participants probably base their judgments on a set of beliefs that make size a more critical feature for defining quarters than pizzas. In such a case, any similarity metric that ignores participants' knowledge that pizza and money are defined by different features will quickly become unable to make any predictions in this domain. Other research extends this critique by showing that the particular context used when making similarity comparisons can dramatically affect the organization of a similarity space. Roth and Shoben (1983) had participants read a sentence that referred to some common category ("the hunter shot at the *bird* flying high overhead") and then verify whether a specific exemplar (for example, *robin*) could have been the referent of the sentence. In this paradigm, high semantic similarity between the exemplar and category term in the sentence speeds verification reaction time (RT). However, Roth and Shoben found that only typicality ratings made in the presence of the specific context provided by the sentence were correlated with verification RT. Contextless typicality ratings (e.g.,

"How typical a bird is a robin?") were completely uncorrelated with verification RT. These data are usually interpreted to mean that conceptual structure is highly constrained by the abstract knowledge participants have about specific domains. These beliefs and theories might serve as organizing principles that help constrain the ways that features are selected when objects are classified, compared, and recognized.

This is not to say that similarity is unimportant. Rather, the issue is what role similarity should play in understanding concepts (see Sloman & Rips, 1998, for a review). One option is to separate a perceptual identification process from a more complex and deliberative classification process, then relegate similarity to the identification function (Armstrong, Gleitman, & Gleitman, 1983). Another alternative that has been recently suggested is to identify a set of broadly applicable task contexts that might predictably constrain similarity (Medin, Goldstone, & Gentner, 1993) or to assume that basic perceptual processes sufficiently constrain similarity (Goldstone, 1994). A third alternative would be to identify a set of particular domains of knowledge and search for regularities in feature selection within those domains (Keil, 1994).

Of these alternatives, the present data suggest that the first would not be helpful. Pushing similarity "down a level" to fast entry-level object classification or identification affords no escape from the problems inherent to its contextual malleability. As noted above, the data presented here seem difficult for a similarity model to adapt to without including just this kind of sensitivity to task context. This is an important point because many models of object recognition proceed as if their emphasis on the "perceptual" process of object recognition allows them to ignore the problems inherent in similarity-based explanations. In the case of faces, all of the research presented in this study and in Levin (1996) test just this type of undeliberated, nonconceptual process; so if the above reasoning is correct, then we cannot separate these tasks from the need for additional explanation.

The second alternative is more promising, although it is not without limitations. According to Medin et al. (1993), similarity models can be constrained by specifying how tasks select features for comparison between objects. This understanding of similarity makes it "less a computation over some feature space than . . . a search process." (Medin et al., p. 261). For example, Medin et al. found that participants segment an ambiguous stimulus in different ways depending on the item it is being compared with. Further, the ambiguity is resolved by having the ambiguous feature take a form that matches the corresponding feature on the context-setting object. Thus, the comparison context both selects the relevant features and influences the form they take. Medin et al. argued that these broadly applicable contextual constraints answer the question "similar in what respect?" Thus, they argued that specifying the "respects" used to make similarity comparisons can make similarity a useful component of a theory explaining human concept formation. Hosie and Milne (1996) made a complementary point about the finding that distinctive faces are more easily recognized. They find that the distinctiveness advantage in recognition is stronger when a mixed set of typical and distinctive faces are presented together, as compared with a design where typical faces are presented in one set and distinctive faces in another. Hosie and Milne suggested that the typical faces set a context that causes more subtle coding of metric differences among faces and therefore allows for improved recognition of distinctive faces. This allows a task context to predictably affect the metrics of a similarity space.

Although specifying task-based constraints on similarity may provide a series of domain-general regularities to apply to object recognition, such constraints seem unable to tap regularities in our use of concepts within specific domains. The present research highlights this limitation. In all experiments, the tasks focused attention on the same set of information: race. For example, in the between-race discrimination trials in Experiments 3 and 4, the difference between the A and B faces focused attention on race-specifying information, thus specifying the particular aspects to be used in the similarity comparison. The methods therefore involved a task that constrained the features used in the comparison in the same way as in Medin et al. (1993) study. Despite the fact that similarity was constrained in the same way for both Black and White continuum end regions, discrimination accuracy was higher for Black faces. This suggests that constraints on similarity need to apply to more than the task context and should also refer directly to more stable characteristics of the participants' conceptual structure.

Accordingly, the domain-general task constraints described above should be considered in concert with more abstract reasoning, knowledge, and even theories (e.g. "mental explanations"; Murphy & Medin, 1985) that constrain and organize representations of specific domains (Keil, 1989; Murphy & Medin, 1985). This understanding of concepts can include the kinds of similarity relationships described above, but it is also organized by more abstract knowledge specific to a given domain. In addition, knowledge can organize perception in some instances, whereas more domain general notions of similarity can take over where domain specific knowledge is not present (Keil, 1994). This approach is different from one in which similarity is confined to perceptual identification because it allows similarity relationships to enter into deliberate conceptualization and further raises the possibility that perceptual identification can be affected by abstract, domain-specific knowledge. It is important to note that domain-specific knowledge can affect first-pass object identification in the absence of complex cognitions by means of cognitive and perceptual schemas that are initially set up in a knowledge-directed learning process.

Accordingly, face categories are based on the social cognitions associated with person classification that cause facial features to be selected using a perceptual schema or frame. This process might be understood in terms of "hard wiring" or perhaps proceduralizing (Anderson, 1983) domain-specific knowledge, but a strong commitment to the specifics of proceduralization would be premature. The key is that, even in apparently perceptual tasks, some account of domain-specific knowledge appears necessary. A good example of this comes from Minsky's (1975) discussion of "room frames" that codify specific perceptual expectations for the structure of stereotypic scenes (e.g., one expects to see a door and perhaps windows in a typical room).

This kind of frame can be seen as a more complex version of global perceptual constraints that appear to be built into the visual system. For example, Von Grunau and Dube (1994) and Kleffner and Ramachandran (1992) found search asymmetries favoring stimuli that stand as exceptions to the prevailing ecology of object perception. In the case of Von Grunau and Dube, uptilted boxes are easier to detect because they are exceptions to the rule that

objects tend to be seen from above; and in the case of Kleffner and Ramachandran, concavities are exceptions to the norm of convexity. If visual search reflects this kind of global perceptual constraint, then constraints that are more specific and perhaps cognitively driven, but nonetheless structurally similar to lower level perceptual constraints, should be reflected in undeliberated object recognition and classification. The feature coding asymmetry in faces can therefore be seen as a constraint on perception that reflects the cognitive and social ecology of face perception. This kind of perceptual–cognitive frame might be seen as a more complex version of basic perceptual constraints on similarity (Goldstone, 1994). However, it operates in an explicitly domain-specific fashion and therefore would be counter to the universalistic nature of similarity-based accounts of object categories.

Conclusion

The feature-selection explanation for the CR recognition deficit emphasizes the importance of understanding why people select some features and not others when recognizing and classifying objects. In doing so, it points up a shortcoming of similarity-based models of objects recognition and classification. The problem is one of focus: Similarity-based models of natural categories look inward at abstract relationships among exemplars to explain variation in classification and identification performance. Although this is a reasonable way of understanding object categories for some purposes, it may fail when one needs to understand natural object categories. In the present case, the most important questions reflect a need to understand who will show the CR recognition deficit and how we can change their behavior. Answering these questions effectively requires moving beyond the assertion that CR faces are difficult to recognize because they are perceptually similar and toward an understanding of the social cognitions and environmental contingencies that influence what specific features people use to recognize and classify faces.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anthony, T., Cooper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin, 18,* 296–301.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13,* 263–308.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281–316.

Bartsch, R. A., & Judd, C. M. (1993). Majority–minority status and perceived in-group variability. *European Journal of Psychology, 23,* 471–483.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217–239.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15,* 19–25.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology, 8,* 306–318.

Carey, S., & Diamond, R. (1977). From piecemeal to configural representation of faces. *Science, 195,* 312–313.

Carroo, A. W. (1986). Other race face recognition: A comparison of Black American and African subjects. *Perceptual and Motor Skills, 62,* 135–138.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A,* 879–894.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10,* 393–396.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Edelman, S. (1995). Representation of similarity in three dimensional object discrimination. *Neural Computation, 7,* 408–423.

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training of the recognition of white and oriental faces. *Bulletin of the Psychonomic Society, 2,* 71–73.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expression. *Cognition, 44,* 227–240.

Fallshore, M. F., & Schooler, J. W. (1995). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1608–1623.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628–634.

Feingold, G. A. (1914). The influence of context on the identification of persons and things. *Journal of Criminal Law and Criminology, 5,* 39–51.

Feinman, S., & Entwistle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506–510.

Flin, R. (1985). Development of face recognition: An encoding switch? *British Journal of Psychology, 76,* 123–134.

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society, 23,* 211–214.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157.

Hosie, J. A., & Milne, A. B. (1996). The effect of experimental design on memory for typical and distinctive faces. *Memory, 4,* 175–197.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97,* 553–562.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Keil, F. C. (1994). Explanation, association, and the acquisition of word meaning. *Lingua, 92,* 169–196.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52,* 18–36.

Kuhl, P. K. (1991). Human adults and infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Lavarkis, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics, 20,* 475–481.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1364–1382.

Levin, D. T., & Beale, J. B. (2000). Categorical perception occurs in newly learned faces, cross-race faces, and inverted faces. *Perception & Psychophysics, 62,* 386–401.

Li, J. C., Dunning, D., & Malpass, R. S. (1994). *Cross-racial identification*

among European-Americans: Basketball fandom and the contact hypothesis. Unpublished manuscript.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61,* 379–388.

Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76,* 587–589.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57,* 165–188.

Malpass, R. S., & Kravitz, J. (1969). Recognition for face of own and other race. *Journal of Personality and Social Psychology, 13,* 330–334.

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

Martini, P., & Nakayama, K. (1998, November). *Spatial summation of facial images reveals configural processing.* Poster session presented at the Object Perpcetion and Memory Conference, Dallas, Texas.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469–1481.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–228.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–280). New York: McGraw-Hill.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 553–557.

Ng, W., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross Cultural Psychology, 25,* 217–232.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

Ostrom, T. M. (1993). Differential processing of in-group and out-group information. *Journal of Personality and Social Psychology, 64,* 21–34.

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

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47,* 25–57.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural encoding in face recognition. *British Journal of Psychology, 80,* 273–283.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosnaidu & A. Ortnoy (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36–71.

Sedikides, C., & Ostrom, T. M. (1993). Perceptions of group variability: Moving from an uncertain crawl to a purposeful stride. *Social Cognition, 11,* 165–174.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

Shepard, J. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepard (Eds.), *Perceiving and remembering faces.* London: Academic Press.

Sloman, S., & Rips, L. (1998). Similarity as an explanitory construct. *Cognition, 65,* 87–101.

Simon, B., & Brown, R. (1987). Perceived intragroup homogeneity in minority–majority contexts. *Journal of Personality and Social Psychology, 53,* 703–711.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory stereotyping. *Journal of Personality and Social Psychology, 36,* 778–793.

Teitelbaum, S., & Geiselman, R. E. (1997). Observer mood and cross-racial recognition of faces. *Journal of Cross Cultural Psychology, 28,* 93–106.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43A,* 4161–204.

Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259–273.

Valentine, T., Chiroro, P., & Dixon, R. (1995). An account of the own-race bias and the contact hypothesis based on a "face space" model of face recognition. In T. Valentine (Ed), *Cognitive and Computational Aspects of Face Recognition: Explorations of Face Space* (pp. 69–94). London: Routledge.

Valentine, T., & Endo, M. (1992). Toward an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44A,* 161–204.

Von Grunau, M., & Dube, S. (1994). Visual search asymmetry for viewing direction. *Perception & Psychophysics, 56,* 211–220.

Yin, R. K. (1969). Looking at upside down faces. *Journal of Experimental Psychology, 81,* 141–145.

Received June 7, 1996
Revision received July 6, 1999
Accepted January 21, 2000 ■