

RACIAL CATEGORIZATION OF FACES The Ambiguous Race Face Effect

Otto H. MacLin and Roy S. Malpass
University of Texas at El Paso

Accusations of discriminatory treatment of minority persons in the criminal justice system create a need for policy and procedure development to create real and perceived equal treatment. A facial recognition deficit among law enforcement officers and witnesses for persons of another “race” contributes to unequal treatment of minority group members. This article demonstrates the other-race effect in an unusual context, reveals theoretical weaknesses, reveals the role of categorical processes in the phenomenon, and discusses policy implications. Experiment 1, based on feature and trait ratings, demonstrates that identical and racially ambiguous faces with different racial markers (hair) are perceived according to the marker. Experiment 2 demonstrates an other-race recognition effect using these faces. A feature acting as a racial marker can cause a face to be perceived and remembered differently. Other-race faces are perceived categorically, which drives the recognition process.

Governor George Ryan of Illinois suspended executions in his state because a large proportion of death row inmates were exonerated and released. Uncertainty as to the actual guilt of death row inmates is a sobering matter and underscores the widely held belief that there are serious problems of accuracy in the criminal justice system. The recent use of DNA evidence to exonerate people who have been falsely incarcerated (Connors, Lundregan, Miller, & McEwan, 1996; Scheck, Neufeld, & Dwyer, 2000) has drawn national attention to the field of eyewitness identification because a majority of those exonerated were convicted based on mistaken identity (Wells et al., 1998). Many of these identifications were other-race¹ identifications, and, as a result, social science information regarding the special difficulties of other-race facial recognition becomes a relevant area for policy consideration.

Eyewitness identification is considered one of the most important methods in apprehending criminals (Lasota, 1974), is considered direct evidence of guilt (Wells et al., 1998), and is accorded a high degree of importance by juries (Loftus, 1975; Penrod & Cutler, 1995). The way in which identification evidence is

Otto H. MacLin and Roy S. Malpass, The Eyewitness Laboratory, Department of Psychology, University of Texas at El Paso.

This research was supported by National Science Foundation Grant SBR-9730937. We thank Dawn McQuiston, Shannon Honaker, Vivian Herrera, and Serena Holguin for their contributions to this research.

Correspondence concerning this article should be addressed to Otto H. MacLin, Department of Psychology, University of Texas, El Paso, Texas 79968. Electronic mail may be sent to omaclin@utep.edu.

¹ We use the term *race* to refer to popular categorizations of persons on the basis of their perceived physical appearance. We do not subscribe to any “scientific” theory of race, and we use the term as a lay or folk concept rather than as a technical term.

gathered and used is the subject of a recent policy document published by the National Institute of Justice (Technical Working Group for Eyewitness Evidence, 1999). These guidelines do not in themselves address the other-race effect and the complexities it embodies. However, a relative deficit among both law enforcement officers and witnesses in facial recognition for persons of another race contributes to unequal treatment of minority group members. The other-race effect has been known through empirical studies since Malpass and Kravitz (1969), and since that time many studies have attempted explanation of the phenomenon (Meissner & Brigham, 2001). Survey results revealed that approximately three-quarters of the experts in eyewitness testimony believed that there existed sufficient and reliable support for the other-race effect that they would personally testify as such (Kassin, Ellsworth, & Smith, 1989). An up-to-date replication and extension (Kassin, Tubb, Hosch, & Memon, in press) finds approximately 90% of experts judging the effect reliable enough for testimony in court. Although the other-race effect is commonly accepted among experts in the areas of law and psychology, there is no widely accepted account of the reason for the effect and the mechanisms through which it works. This article attempts to increase understanding of the other-race effect in perception and recognition and to identify implications for policy and procedure.

The other-race effect is not an isolated effect. Many of us have experienced the feeling that people of other races appear more similar to each other than people of our own race. The statement "They all look alike to me" sums up the notion that other-race faces seem to show less variation among individuals than own-race faces. Quite beyond the common occurrence of the effect, however, are events that put people in prison based on errors of identification. Ronald Cotton, a Black man, was convicted of raping 22-year-old White college student Jennifer Thompson based on her eyewitness testimony. After 11 years in prison he was exonerated by DNA evidence. During the rape, Jennifer Thompson had studied the face of her attacker. She was certain it was Mr. Cotton that had raped her, but she was wrong (Thompson, 2000).

The other-race effect was first demonstrated by Malpass and Kravitz (1969) using a yes–no recognition task, and since that time the phenomenon has drawn much research interest and has been the subject of several meta-analyses (Anthony, Copper, & Mullin, 1992; Bothwell, Brigham, & Malpass, 1989; Meissner & Brigham, 2001; Shapiro & Penrod, 1986) and other reviews (Chance & Goldstein, 1996). Despite the ease with which the other-race effect is demonstrated psychologists have yet to provide a convincing explanation for this widely replicated phenomenon. Two popular explanations for the other-race effect, social attitudes and differential contact, are often mentioned, but clear support is seldom provided for either. Research has failed to produce consistent and reliable empirical evidence to support the social attitudes hypothesis (Brigham & Barkowitz, 1978; Lavrakas, Buri, & Mayzner, 1976; Malpass, 1992; Meissner & Brigham, 2001; Platz & Hosch, 1988). For example, self-reported interracial contact between White students from Canada and Asian students from Singapore was not significantly related to recognition performance of other-race faces (Ng & Lindsay, 1994).

The amount of contact individuals have with other races is also said to be related to recognition of other-race faces, but the evidence is sporadic, and process

theories have not been accompanied by supporting data (Brigham & Malpass, 1985; Malpass, 1990, 1992). A number of studies have shown general support for the differential experience hypothesis, which suggests that the ability to recognize faces of another race is a function not of the absolute amount of contact one has had with members of that race, but the quality of the contact (Brigham, Maass, Snyder, & Spaulding, 1982; Brigham & Malpass, 1985; Carroo, 1986; Lavrakas, Buri, & Mayzner, 1976; Platz & Hosch, 1988). Face recognition skills may develop from a need to individuate members of populations who are important to be able to identify, such as parents, bosses, and other influential social contacts (Malpass, 1990, 1992). Frequent contact with members of another race should not increase recognition ability if there is no need to be able to identify and recognize individuals of that race. This hypothesis received some support from Platz and Hosch (1988) and can be thought of as related to Nickerson and Adams's (1979) famous demonstration that although most people can readily distinguish pennies from other coins, very few can recognize the correct features of the penny in a penny "lineup" where the correct pattern of features is displayed among others with incorrect patterns. Color and size are all that is needed to distinguish pennies from other coins. It is rarely important to be able to identify individual pennies, so the particular arrangement of features (e.g., which way Lincoln is looking) on pennies goes unlearned. The extrapolation to other-race recognition is direct. Features that mark group membership are all that is needed to categorize people into "own" and "other" categories. The particular features that differentiate individuals within any group will be learned to the extent that it is important to differentiate between individuals in the category in the course of everyday life. A good example of this is the superior recognition for Black faces by White basketball fans in Dunning, Li, and Malpass (1998).

Goldstein (1979a) attributed the difficulty in recognizing other-race faces to a lack of expertise with the feature dimensions on which other-race faces vary. The concept of perceptual learning has been proposed as an important mechanism for understanding how differential experience affects the way we process own-race and other-race faces (O'Toole, Abdi, Deffenbacher, & Valentine, 1995). According to the differential experience hypothesis, as face recognition skills develop, individuals learn to use the perceptual dimensions that are optimal for discriminating among individual faces. However, the perceptual dimensions on which own-race faces vary are not always the same for different races, because participants differ in their descriptions of same- and other-race faces (Ellis, Deregowski, & Shepherd, 1975), ratings of distinctiveness (van Wallendael & Kuhn, 1997), and attractiveness (Milord, 1978). These perceptual differences are not a result of a lack of variability within each race (Goldstein, 1979a, 1979b). Typically, because individuals have more experience with own-race faces, they become more familiar with the feature dimensions that are most salient for distinguishing own-race faces. Perceptual learning greatly enhances individuals' ability to process and recognize own-race faces, but it also results in a decreased ability to recognize other-race faces.

In an attempt to understand how people recognize faces, Valentine (1991) proposed the concept of a multidimensional "face space," which in part is based on perceptual learning as described above. Valentine's face space model attempts to provide an explanation of how all of the faces a person has ever encountered

are stored in memory as mental representations. One way to help conceptualize the face space model is by making a comparison with another perceptual phenomenon based on color vision: the "color space." Color space can be thought of as a three-dimensional model consisting of red, blue, and green dimensions (RGB). Depending on the contribution of the individual RGB values, any color in the environment can be represented, or located within color space. As an example, the pixel values on a computer monitor or color television are based on RGB color space, and depending on the combination of RGB values specified by a program a particular color will be displayed. Inversely, if a color is known, the RGB values can be derived.

According to the face space model, as with the color space model, individual faces are represented as combined locations on relevant perceptual dimensions. Unlike the color space, the dimensions of face space have not been entirely worked out; however, psychologists have indications of what some of those dimensions might possibly be. Ellis, Deregowski, and Shepherd (1975) found that European Whites referred more to hair, hair texture, and eye color when describing faces, whereas Black Africans referred to features such as eye size, eyebrows, and ears. Additionally, when female faces were sorted for attractiveness by Black participants and White participants, it was found that the strategy used to sort a mixed set of faces was primarily based on race of the face, size and shape of nose, thickness of lips, hair color, and skin color and texture (Shepherd & Deregowski, 1981). Sergent (1984) collected similarity judgments for pairs of faces developed using a Photo-fit composite program. Sergent found that faces are also represented by their individual configuration, not just their individual features. Similarly, Haig (1984) found that participants were sensitive to small changes in the distance between the mouth and nose and the separation between the eyes, again indicating that some of the dimensions of face space are likely based on the configural properties of the face and not just the features. Finally, there is evidence that face space may have properties similar to color space such as neutral center point, equivalent to that of the color white in color space, which would be said to represent typical faces in face space (MacLin & Webster, in press; Webster & MacLin, 1999).

The dimensions of face space are more complex and less well understood than those of color space. To simplify the discussion of face space, we refer to face space as an n -dimensional model compared with the three-dimensional model of color space. Although the dimensions of color space are largely dictated by the constraints of our visual system, such as the color receptors in the retina, it is believed that the dimensions of face space are developed as a product of all of the faces a person has ever encountered in their lifetime. This is consistent with the perceptual learning theories.

In face-recognition tasks it has been demonstrated that distinct faces are easier to recognize than are typical faces (Chiroro & Valentine, 1995; Shepherd, Gibling, & Ellis, 1991; Valentine & Bruce, 1986; Valentine & Endo, 1992), and the face space model has been able to account for the effects of distinctiveness (Valentine, 1991). For example, when a face is encountered, it acts as a memory probe. If the probe is from a typical face, similar faces will be located in a densely populated region of face space causing many "faces" to be activated (more accurately, "face representations," because we do not know precisely how faces are represented cognitively). With many faces activated, it is difficult to discriminate among the many faces resembling the probe. This in turn causes a higher

occurrence of recognition errors. If the probe is a distinctive face, faces located in a sparsely populated area of face space will be activated. There will be fewer stored face representations to match with the probe because there will be fewer faces located in that area. With fewer face representations activated, recognition performance will be more accurate with distinct faces.

There are two models explaining how faces are encoded in face space: *norm-based* and *exemplar-based*. It is not yet known whether faces are encoded based on their relationship to just one (norm-based) central prototype or relative to each of the other faces distributed within the (exemplar-based) face space (Valentine, 1991; Valentine & Endo, 1992). Both models do assume that all faces that an individual has ever encountered are represented in the face space. Both models make the same predictions about recognizing distinct faces and faces of another race.

Figure 1 demonstrates how norm-based encoding is represented as feature vectors. In the norm-based model, faces are encoded based on their similarity to the central prototype, which acts as an abstraction of all the faces stored in the face space (see Goldstein & Chance, 1980), where faces that are less similar to the prototype are located further away from the prototype and have longer vectors, whereas faces similar to the prototype are located near the center of the face space and have shorter vectors (as illustrated on the left side of Figure 1).

In contrast, exemplar-based encoding is represented as discrete points located within the face space. These points are encoded relative to all of the other faces in the face space, not to a central prototype. With the exemplar-based model, each new face is compared with all of the other faces located in the face space. The new face is then located such that it is closest to the faces with the greatest resem-

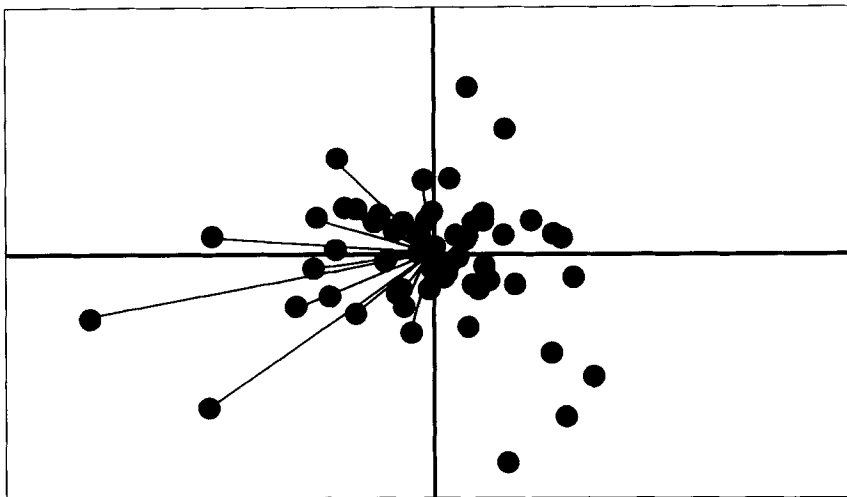


Figure 1. Multidimensional face space. Faces to the left on Figure 1 represent norm-based encoding where faces are encoded relative to the prototype located in the center of the face space. Connecting lines represent vectors indicating face is compared to central prototype. Faces on the right side of Figure 1 represent exemplar-based encoding where each face is encoded relative to the other faces located in face space.

blance, thus sharing similar perceptual dimension values as illustrated on the right side of Figure 1.

How does the face space model explain the other-race effect? As previously mentioned, it is hypothesized that the dimensions underlying face space are determined by the perceptually salient features of all faces ever encountered (Valentine, 1991). Therefore, these dimensions will be primarily based on (specialized to) own-race faces because they are encountered more often, and there is thus a greater need to differentiate among own-race individuals. In contrast, the dimensions for other-race faces, which are encountered less often, will be ill-defined as they will be encoded based on same-race dimensions, causing faces of other races to be represented in dense clusters and located away from the center of the face space (see Figure 2). Although other-race faces will be located a distance away from the center of the face space, they will also be more difficult to recognize because of the density of their distribution in the space (Valentine & Endo, 1992).

Recently face space models have been tested using caricatures of own and other race faces to determine if the norm-based or the exemplar-based model best describes how faces are encoded. Byatt and Rhodes (1998) demonstrated that faces are encoded not merely as deviations from a norm but as discrete points in face space (exemplar-based). Byatt and Rhodes used a caricature generator (Brennen, 1985; Rhodes, Brennen, & Carey, 1987) to create four sets of faces based on deviations from a prototypical face. These four sets were either based on deviations from a same-race prototype (right-norm distortions) or an other-race prototype (wrong-norm distortions). Specifically, two sets of European caricatures



Figure 2. Faces are clustered more densely in the center of the face space representing typical same-race faces, whereas faces in the lower right quadrant (circled for demonstration purposes) representing other-race faces are densely clustered as well. Faces that are similar with the exception of one feature (e.g., nose) are located near each other, as for Faces 1a and 1b. However, when a key feature acting as a racial marker (e.g., hair) is changed, the face with a racial marker of another race is located in the dense cluster representing other-race faces, as in Faces 1a and 2a.

were created with either a European prototype (right-norm) or a Chinese prototype (wrong-norm), and two sets of Chinese caricatures were created using the Chinese prototype (right-norm) or a European prototype (wrong-norm). The norm-based model predicts that European participants are better able to recognize caricatures based on the European prototype (right-norm European and wrong-norm Chinese faces) regardless of the race of the caricature faces. On the other hand, because the exemplar-based model is not derived from a central prototype, it predicts that right-norm faces will be better recognized than wrong-norm faces. Byatt and Rhodes (1998) found that right-norm caricatures were better recognized than European norm-based caricatures, supporting the exemplar-based face space model over the norm-based model.

The present study further investigated how the face space model accounts for the other-race effect by using ambiguous race faces, where otherwise identical faces are transformed from one race to another by changing one feature that functions as a racial marker. Racially ambiguous faces were created such that certain facial features (e.g., eyes, nose, and mouth) overlapped across Hispanic and African American racial lines. Because the features overlap, the racial marker feature is absent, and the face is ambiguous as to which race it belongs. When the racial marker feature is added, the ambiguity is resolved, and the face is readily perceived as a member of a particular racial category. Hair was used as the racial marker for the faces in this study. Ambiguous race faces were created absent the hair feature.

According to the multidimensional face space model (Valentine, 1991), similar faces should be located near each other, and they should be equally recognizable because they will be located in an area of face space with equal density, as demonstrated in Figure 2 with Faces 1a and 1b. However, if the face is perceived as an other-race face, it should be located in an area separate from own race faces and be more difficult to recognize than own-race faces because the area of face space representing other-race faces is more densely populated, as shown in Figure 2 with Faces 1a and 2a. If otherwise identical, ambiguous race faces are perceived and categorized differently because of the racial marker. If an other-race deficit can be demonstrated in a face-recognition task, both exemplar- and norm-based models would have difficulty accounting for the result, because they are both based on the similarity of the faces represented.

In Experiment 1 we examined the ambiguous race face effect by determining if the faces were perceived as either Black or Hispanic depending on the racial marker. In Experiment 2, we examined whether an other-race effect on recognition was obtained using ambiguous race faces.

Experiment 1

Method

Participants. Twenty students (5 men, 15 women) from the University of Texas at El Paso (UTEP) participated in this experiment for partial course credit.² UTEP is located on the United States–Mexico border, and the Hispanic students at UTEP account for

²Participants in this study were treated according to the American Psychological Association Ethical Principles of Psychologists and Code of Conduct (1992).

approximately 68% of the student population compared with 17% White and 2% Black students. Additionally, 9% of the students at UTEP are Mexican nationals. Participants in this study were self-identified as "Hispanic" by self-report in a free response format, while signing in for the experiment. For the purposes of this study no attempt was made to differentiate the nationality of the Hispanic participants. We also note that even though participants from both racial populations are normally used in the collection of other-race data so that the existence of a crossover effect can be determined, this was not a consideration in this experiment, because the stimuli used in each racial category consisted of exactly the same face, with hairstyle as the single differentiating feature.

Stimuli. Ambiguous race faces with an Hispanic–Black feature overlap were created using a facial composite construction kit (Faces 3.0), such that the facial features were general (ears and chin) or overlapping across racial lines (dark eyebrows, dark eyes, broad nose, and full lips³; see Figure 3A). When a key feature characteristic of a particular race is added to the face, in this case hair, it acts as a racial marker, and the face is perceived as a person of that race (Figures 3B and 3C). Based on pilot work, hairstyles were chosen on the basis of their stereotypicality for each of the two races. The particular ambiguous race faces used in this study were selected from a larger set of 150 faces based on ratings from five independent Hispanic judges. The 20 ambiguous race faces that were classified most often according to their racial marker were chosen for this experiment. From these 20 faces, 20 Hispanic faces were created using a Hispanic hairstyle (Figure 3B), and 20 Black faces were created by using an Afro hairstyle (Figure 3C). Thus, each participant saw a total of 40 faces. In order to reduce any potential carryover effect in Experiment 1, faces were presented in two blocks of 20, where each ambiguous race face (with racial marker) was presented only once in each block. If a face was presented with an Hispanic hairstyle in the first block, it was presented with an African American hairstyle in the second block, and vice-versa.

Procedure. Face images were set at 248×232 pixels and displayed one at a time on a computer monitor set at 800×600 pixels. Participants were shown a face on the monitor and asked to categorize the race as "Indian," "Asian," "Black," "Hispanic," "White," or "Other" by clicking on the corresponding word located on the screen below the face image. Once the face was classified for race, participants were asked to rate the degree to which the face fit that racial category on a 9-point Likert-type scale. For example, if the participant categorized the face as Black, the anchors read "Not at all Black" and "Very Black." Participants indicated their response by clicking on one of the nine option boxes located on the screen directly below the facial image and between the corresponding anchors. Numbers in the scale were not visible to the participants. Following the raceness rating, participants made 11 ratings of facial features and 7 ratings of personality traits (see Appendix) that appeared in the center of the monitor in a random order. The classification and rating task required approximately 50 min. Upon completion of the experiment, participants were debriefed as to the nature of the experiment, thanked for their participation, and dismissed.

Results

Overall, 68% of the Black faces (ambiguous race face + Afro hair feature) were classified as Black, 7% as Hispanic, 1% Indian, 3% White, 2% Asian, and

³Research with a group of 103 students comparing stereotypes of facial features between Blacks and Whites found that 95% of the participants indicated that Blacks have larger lips than Whites, 86% indicated that Blacks would have "kinky" hair more so than Whites, and 90% indicated that Blacks were more likely to have broader noses than Whites (Swim & Shotland, 1999).

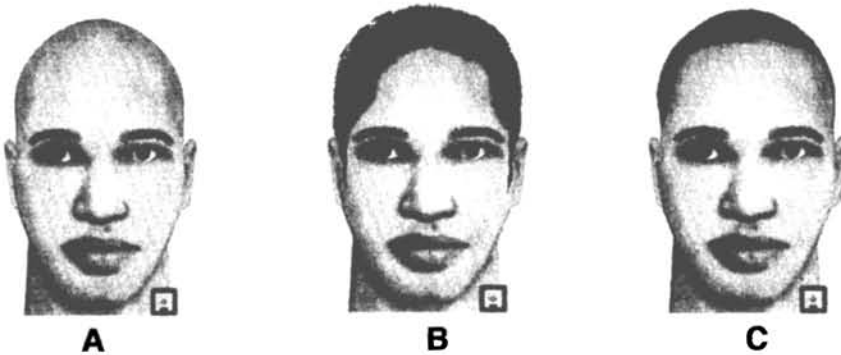


Figure 3. Ambiguous race faces. A represents a unique face (without key feature hair); B represents a unique face + Hispanic hairstyle; and C, a unique face + African American hairstyle.

19% as Other. Of the Hispanic faces (ambiguous race face + Hispanic hair feature), 68% were classified as Hispanic, 1% as Black, 7% Indian, 3% White, 3% Asian, and 18% Other (percentages were rounded; see Figure 4). Faces classified as Hispanic were rated higher on raceness than faces classified as Black, $t(19) = 3.92, p < .001$. Given that approximately two thirds of the faces were perceived according to the key feature, subsequent analysis is limited to the faces categorized according to their racial marker.

Rating scores for faces classified as either Black or Hispanic were averaged across participants for each of the 40 faces for raceness, the 11 features, and the 7 personality attributes. The means of these attributes for the 20 Hispanic faces were compared with the means of the 20 corresponding Black faces using a repeated measure t test. Black faces were perceived to have a darker complexion ($M = 4.42$) than Hispanic faces ($M = 4.07$; see Figure 5), $t(19) = 2.48, p < .05$. The mouths on the Black faces were perceived as wider ($M = 5.58$) than mouths on the Hispanic faces ($M = 5.88$), $t(19) = 2.24, p < .05$. Hispanic eyes were rated as more protruding ($M = 3.08$) than Black eyes ($M = 2.82$), $t(19) = 2.16, p < .05$. Black faces were rated as wider ($M = 4.81$) than Hispanic faces ($M = 4.46$), $t(19) = 2.56, p < .05$.

The personality of Hispanic faces was rated as "stronger" than that of Black faces ($M = 4.92$ and $M = 4.70$, respectively), $t(19) = 2.14, p < .05$; more "submissive" ($M = 3.56$ and $M = 3.16$, respectively), $t(19) = 2.17, p < .05$; more "tense" ($M = 4.43$ and $M = 3.96$, respectively), $t(19) = 2.34, p < .05$; and more "suspicious" ($M = 4.93$ and $M = 4.28$, respectively), $t(19) = 4.41, p < .01$. Lastly, Black faces were perceived as "warmer" than Hispanic faces ($M = 3.64$ and $M = 3.31$, respectively), $t(19) = 2.40, p < .05$.

Discussion

There was no difference between the number of faces classified as Black or Hispanic, indicating that the selection of the ambiguous race face stimuli showed no effect toward either race. Slightly more than two thirds of the faces were classified according to their racial marker. This suggests that the majority of the

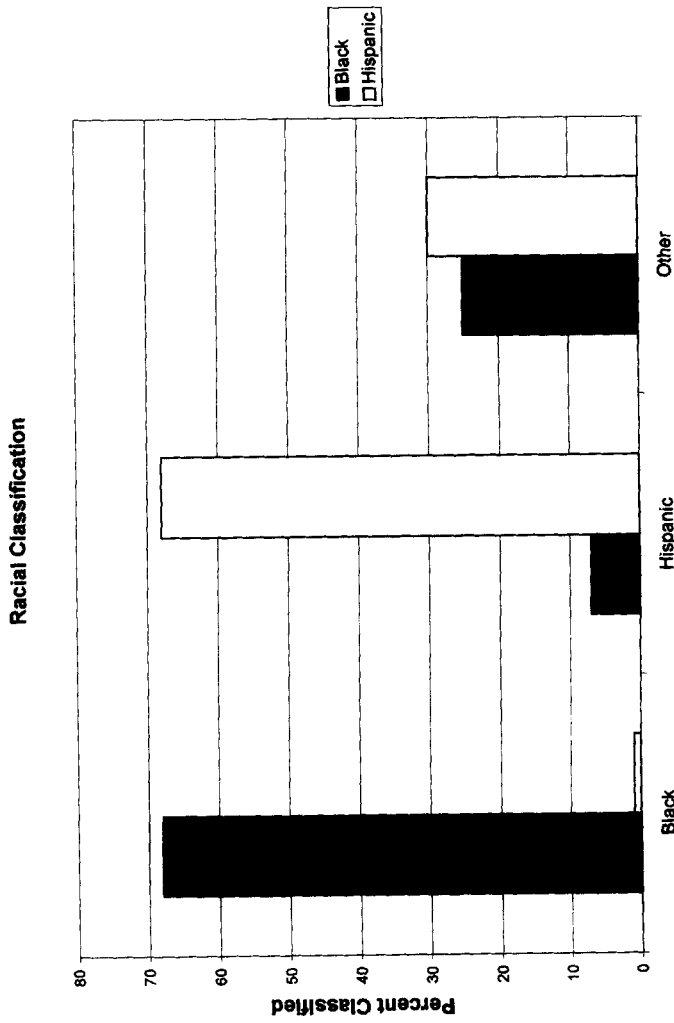


Figure 4. Results from racial classification task. Results for classification other than Black and Hispanic are represented as "Other."

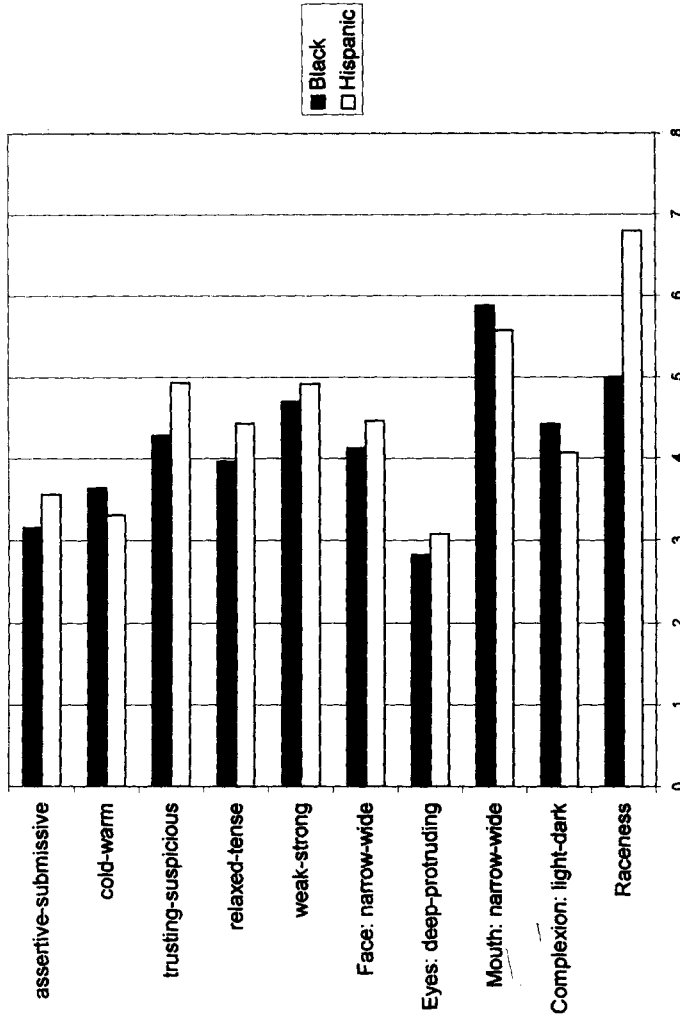


Figure 5. Results from feature- and trait-rating task ($p < .05$).

faces were truly racially ambiguous, so that the hair feature indeed acted as a racial marker and had a strong influence on the categorization and perception of race. A higher rate of classification would probably have been obtained if the participants were constrained to select among only two racial categories (Hispanic or Black).

Faces perceived as Hispanic were rated as more Hispanic on the raceness scale than those faces that were perceived as Black. There are several potential explanations for this outcome. It may result from limitations of the composite program: Feature representations for Hispanic faces might be more effectively implemented than for African American faces. Alternatively, the Hispanic faces, composed by Hispanic and White researchers at a predominately Hispanic university, may have been better representatives of Hispanic faces, based on the composers' expertise with this category of faces. Finally, it is possible that once a face is classified as own-race, there is social value in rating it higher in raceness (Blascovich, Wyer, Swart, & Kibler, 1997).

If these identical faces (with the exception of one feature) were closely located in face space, as would be expected by the face space model, one would not predict that they would produce such variable feature and personality judgments. It appears that the racial marker drives racial categorization that in turn drives the perceptual process. It is possible that the participants are judging the faces based on social or stereotypical criteria. Alternatively, it may be that the mental representations of the two similar faces are stored in completely different locations of face space, one representing same-race and the other representing other-race faces. If this is true, then the feature and personality judgments are relative not to similar faces, but relative to nearby faces stored within that particular area of face space representing the racial category indicated by the racial marker.

Experiment 2 was designed to determine whether the other-race effect occurs using the ambiguous race face stimuli in a recognition task. If the other-race effect can be demonstrated using ambiguous race faces, this would controvert the differential experience hypothesis, which states that the other-race effect is a result of encoding other-race faces on own-race dimensions. Because there is no physical difference between the two stimulus faces other than the racial marker, participants should extract and encode the same individuating information for each face whether it has Hispanic or African American hair, and this information should be equally useful in recognizing the faces in a subsequent recognition test.

Experiment 2

Method

Participants. Fifteen Hispanic participants from UTEP participated in this experiment for partial course credit. Participants were treated according to the ethical standards as outlined by the American Psychological Association.

Stimuli. Twenty-four ambiguous race faces from the set of 40 used in Experiment 1 were selected. Rather than using only one hairstyle for each race, four hairstyles from a larger pool of 12 hairstyles were chosen for each race based on ratings from independent judges. For the Hispanic style hair, 12 Hispanic judges rated 15 hairstyles on stereotypicality, and for the Afro style hair, 12 Black judges rated 15 hairstyles based on stereo-

typicality. Twenty-four "Hispanic" faces, 6 with each hairstyle, and 24 "Black" faces, 6 with each hairstyle, were created for the 24 ambiguous race faces. Half of the faces for each race, three of each hairstyle, were selected for the training phase. The remaining half were introduced as new faces in the recognition phase. Presentation of the faces was counterbalanced, so that half of the time faces were presented as "old" and half of the time as "new."

Procedure. Participants were informed that they would be shown some faces on the computer monitor and then asked questions about them at a later point in the experiment. During the training phase, faces were presented for 3 s with a 5-s interstimulus interval. Faces were presented randomly with the restriction that no three faces of the same race were shown in sequence. Once all 24 old faces were presented, participants were given a 5-min filler task prior to the recognition phase that involved finding hidden words within a matrix of letters. During the recognition phase, the 24 new faces were combined with the 24 old faces (half "Hispanic" and half "Black") and presented one at a time in random order for 5 s with an 8-s interstimulus interval. Participants were instructed that half of the faces had been seen previously and to press the "left arrow" key to indicate if the face was old and had been seen before or to press the "right arrow" key if the face was new and had not been seen before. The use of arrow keys was counterbalanced such that half of the participants pressed the left arrow key for old, and half pressed the left arrow key for new. Again, the restriction that no three faces of the same race and no three old or new faces be shown in sequence was used.

Results

A' is a nonparametric measure of recognition accuracy derived from signal detection theory (Rae, 1976). The advantage of using A' as a measure of recognition accuracy is that it corrects the correct identification rate (hits: saying a face was seen before when it was in fact seen before) for the rate at which the observer said a face was seen before when it actually was not (false alarms: guessing rate). In that way A' provides an estimate of recognition accuracy that is uncontaminated by the observer's response bias. Misses and correct rejections are complements of hits and false alarms and therefore are ignored in the analysis. Higher A' values are associated with increased recognition accuracy. Measures derived from signal detection theory have been used in other-race facial recognition studies since the first empirical study on the topic (Malpass & Kravitz, 1969).

A' scores were higher for ambiguous race faces with Hispanic hairstyles, indicating that Hispanic faces were recognized better than Black faces by the Hispanic participants ($M = .705$ and $M = .576$, respectively), $t(14) = -2.24$, $p < .05$ (see Table 2). A' scores for recognition are shown in Table 1. The difference

Table 1
Possible Outcomes of an "Old"–"New" Recognition Task

Recognition stimulus	Participant's response	
	"Old" response	"New" response
"Old" (seen previously)	Correct I.D. (Hit)	False rejection (1 – Hits)
"New" (not seen previously)	False alarm (FA)	Correct rejection (1 – FA)

Note. I.D. = identification.

in A' scores is largely due to a disproportionate amount of false alarms in recognizing the Black faces. These results are consistent with findings from other other-race research (e.g., Bothwell, Brigham, & Malpass, 1989; Platz & Hosch, 1988; Shapiro & Penrod, 1986).

Discussion

Results from this study indicated that identical faces are perceived and recognized differentially when a key feature acts as a racial marker. This suggests that the other-race face deficit is not necessarily caused by inexperience with the other-race stimulus class as the perceptual learning hypothesis would predict, but by the perceptual categorization of the race. Although this finding is contrary to the predictions of the face space models, it is consistent with the literature on perceptual categorization of faces (Beale & Keil, 1995), facial expressions (de Gelder, Tenuisse, & Benson, 1997), speech (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970), graphemes (Yasuhara & Kuklinski, 1979), and color (Livingston, Andrews, & Harnad, 1998).

Current face space models have difficulty explaining data from ambiguous race faces because these models hold that similarities among faces are based on perceptual dimensions developed from experience with the larger stimulus class, rather than separately, upon categorical subsets. Regardless of whether faces are encoded relative to the exemplars or relative to a norm, identical and nearly identical faces should be encoded very close together in the face space, based on their similarity, just as colors with similar RGB values should be located near each other. Identical faces should be located near each other in areas of face space with equal density. And if the faces are located in close proximity, the face space models would not predict differential recognition as reported in Experiment 2. The results for ambiguous race faces do not fit this model. The two experiments presented here indicate that at least some faces are coded categorically. To allow for categorical perception, face space models may have to be modified to account for an other-race/own-race dimension or a categorization step in their associated process models.

Levin (in press) examined the other-race effect using a visual search task. The visual search task has been used to examine low-level visual feature processing such as color (Carter, 1982), orientation (Cavanagh, Arguin, & Treisman, 1990) and motion (von Gruenau, Dube, & Kwas, 1996). For example, when a feature such as a circle colored red is presented in a group of circles colored blue, it will pop out of the background and be attended to immediately. If red was not a preattentive feature, it would take longer to find the red circle because a serial search would be required. Although Levin did not find a “pop out” effect for

Table 2
Results From Recognition Task

Race of face	Hits	False alarms	A-prime
Black	.57	.49	.576*
Hispanic	.56	.22	.705*

* $p < .05$.

Black faces, he did find a search asymmetry in that a Black face among distracters consisting of White faces was detected faster than a White face among Black faces. A search asymmetry suggests that race may be processed as a higher-order feature. If race is encoded as a feature, it is possible that race may be one of the dimensions in face space.

Possibly, Levin (in press) did not find a search asymmetry with White faces because White acts as a "cultural default" (Zárate & Smith, 1990) and is inferred in the description of individuals. However, if a person is not of the default race (White) the difference is made explicit. For example in describing a tall, White male in his 40s, White is assumed and omitted from the description. However, if the individual is Black and Black is omitted from the description, Black would not necessarily be assumed.

In face space, it is likely that own-race is the default, and race is not encoded as it comprises redundant information. Because race is informative, it is encoded as an additional dimension in face space, however race doesn't help to individuate the exemplar faces. When it is necessary to individuate members of another race, race provides little or no information and may be the source of the higher number of false recognitions for Black faces.

General Discussion

Identical faces can be perceived as faces of different races when a feature acting as a racial marker is changed to that of another race. This presents a problem for the differential experience model that predicts that faces varying on the same dimension should be recognized equally well. Additionally, the face space model predicts that similar faces will be located in the same proximity and that other-race faces will form their own central tendency in face space because they are encoded on poorly defined featural dimensions. Ambiguous race faces are very similar (see Figure 3) and should be located closely in a face space (see Figure 1), but they are not. Our data suggest that other-race faces are encoded categorically and that categorization drives the perceptual process.

Future research with ambiguous race faces should develop additional face stimuli so that a greater percentage of faces are classified according to the racial marker. Additional features and racial markers should be examined to determine if other racial dimensions such as Hispanic-Asian (e.g., using the eyes as a racial marker) also precipitate the other-race effect. Studies should also be symmetrical regarding observers and faces. In the present study, data from Black participants would be required to determine if a crossover effect exists. Furthermore, data from a speeded classification task are needed to determine if reaction times to classify other-race faces are shorter than the time to classify own-race faces (Valentine & Endo, 1992).

False-positive identification of faces not previously seen was the main contributor to the lowered recognition scores in Experiment 2. As seen in the discussion of face space models and the important role that racial categorization plays in the present research, social processes that lead to the racial categorization of persons by perceivers or witnesses can be expected to continue to contribute error even in the concrete area of other-race facial recognition and identification.

The problematic effects of racial categorization are not unique to eyewitness identification, however. They exist in many areas of law enforcement and in many other areas of public life in the United States and other nations.

In a recent news report, Attorney General Janet Reno, commenting on the disproportionate number of minorities on death row in the federal system, said that this was a reflection of the social conditions and opportunities that disproportionately effect minorities (Frieden, September 12, 2000). In the present social climate it may be difficult to ameliorate the effects of racial categorization and other aspects of the bias toward better encoding and recognition of other-race faces. This is mostly because the process begins at the encoding stage of memory, where the criminal justice system has minimal control and impact. Further into the process of identification, greater control is in the hands of the criminal justice system, but the aspects that can be impacted are less important. More care can be taken with eliciting information from witnesses (Fisher & Geiselman, 1992; Malpass, 1996) as indicated by the Technical Working Group on Eyewitness Evidence (1999), and greater care can be taken to follow existing guidelines for constructing lineups, especially to ensure that criminal suspects do not stand out (Technical Working Group on Eyewitness Evidence, 1999).

From a procedural perspective, at the level of lineup construction we should assume race does not exist. This is recommended because categorization drives encoding and recognition processes, and because racial categorization influences the verbal description process. The witnesses should not be induced to make a racial categorization, as its most likely effect will be to aggregate the remembered person more closely with others who share the racial category membership. This will result in making it less likely that the witness will be able to distinguish him from others and more likely that he or she will confuse (and erroneously identify) someone else as being the perpetrator. Furthermore, verbal descriptions should be elicited in an open-ended manner, with follow-up requests for elaboration. Again, race should not be asked about, but if race comes out of the witness's descriptions, it should be accepted without emphasis.

Lineups should be constructed using explicit techniques as outlined by the Technical Working Group on Eyewitness Evidence (1999). Fillers should not be chosen at the level of racial category. Rather, they should be chosen at the feature level, using similarity judgments that refer to facial information. If facial information is used, one does not have to introduce race to make similarity judgments. Depending on the population context, the candidates for lineup fillers should not be restricted to one race so long as the physical features match, even though race may have been indicated by the witness.

It is a fact of life in many urban environments that persons belong to multiple ethnic, identity, and appearance groups. A person could be described as "Hispanic," "Black," "Mexican," "Indian," or "Asian," and have some facial features stereotypically characteristic of other groups. To accept the witness's racial categorization of a perpetrator in such areas may lead to lineups that miss important feature pools on which the choice of lineup fillers should draw.

At the level of the courts, our research shows that we are still learning about the causative social and cognitive factors on which the other-race effect depends. We have also learned that previous assumptions about its cause have taken on

greater complexity and that some explanations discussed in recent decisions (e.g., *State v. Cromedy*, 1999) suffer from a lack of evidence. For example, we have long known what Meissner and Brigham (2001) have shown—that there is a negligible relationship between social attitudes and other-race face recognition and that the theoretical understanding of this relationship is unclear. Likewise, there is only a very small relationship between self-rated other-race social experience and other-race face recognition. This finding has been observed consistently since the first empirical study of the effect (Malpass & Kravitz, 1969). The available evidence points to the role of functional experience that gives importance to distinguishing other-race individuals from one another. But there are no known measurement instruments with which to characterize social interaction environments that might produce such experience, or to measure the degree to which any individual has been exposed to such an environment. We do know that merely being in a place where a lot of other-race individuals are viewed or encountered does not provide a sufficient basis for expertise in other-race face recognition. We know that the other-race effect on face recognition exists, but we do not know what basis might exist for predicting that a specific person might be subject to it or exempt from it.

Even though the basis for the cross-race effect is not fully understood, this is not justification for dismissal of the phenomenon all together. Great care should be used in cases where the witness and suspect are of different race, and eyewitness identification is the primary evidence.

Beyond procedural care in the areas where investigators have control of the process, the use of expert witnesses to inform juries about the increased likelihood of error in other-race identifications may be useful and has been recommended by others (Meissner & Brigham, 2001). At a minimum, cautionary instructions about the influence of other-race identifications should be given to the jury, especially in the case where the evidence consists only of eyewitness identification (e.g., *United States v. Telfaire*, 1972). However, the findings of Devenport, Penrod, and Cutler (1997) concerning the effectiveness of safeguards and the knowledge of jurors on this and many other matters does not inspire confidence in our ability to ameliorate other-race identification errors through informing the finders of fact. Complete elimination of other-race identification errors may have to wait to be overtaken by developments in society at large that deemphasize categorical contrasts or that provide for more equal frequency and importance of social interaction with members of the various visually identifiable categories of persons. In the meantime, careful application of investigative and identification processes according to available good practice standards can reduce the frequency error.

References

- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, *47*, 1597–1611.
- Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality & Social Psychology Bulletin*, *18*, 296–301.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

- Blascovich, J., Wyer, N. A., Swart, L. A., & Kibler, J. L. (1997). Racism and racial categorization. *Journal of Personality and Social Psychology*, *72*, 1364–1372.
- Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin*, *15*, 19–25.
- Brennen, S. E. (1985). The caricature generator. *Leonardo*, *18*, 170–178.
- 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.
- Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). The accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology*, *42*, 673–681.
- Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-races persons. *Journal of Social Issues*, *41*, 139–155.
- Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*, 2455–2486.
- Carroo, A. W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual & Motor Skills*, *62*, 135–138.
- Carter, R. C. (1982). Visual search with color. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 127–136.
- Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 479–491.
- Chance, J. E., & Goldstein, A. G. (1996). The other race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 153–176). Mahwah, NJ: Erlbaum.
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology*, *48A*, 879–894.
- Connors, E., Lundregan, T., Miller, N., & McEwan, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Alexandria, VA: National Institute of Justice.
- de Gelder, B., Teunisse, J., & Benson, P. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition & Emotion*, *11*, 1–23.
- Devenport, J. L., Penrod, S. D., & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluating commonsense evaluations. *Psychology, Public Policy & Law*, *3*, 338–361.
- Dunning, D., Li, J., & Malpass, R. S. (1998, March). *Basketball fandom and cross-race identification among European-Americans: Another look at the contact hypothesis*. Paper presented at the biennial conference of the American Psychology–Law Society, Redondo Beach, CA.
- Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Description of white and black faces by white and black subjects. *International Journal of Psychology*, *10*, 119–123.
- Faces: The ultimate composite picture (English Version 3.0) [Computer software]. (1998). Quebec, Canada: InterQuest Inc.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing*. Springfield, IL: Charles Thomas.
- Frieden, T. (2000). Reno says she finds no basis for moratorium on federal death penalty. *CNN*, September 12, 2000.
- Goldstein, A. G. (1979a). Facial variation: Anthropometric data II. *Bulletin of the Psychonomic Society*, *13*, 191–193.

- Goldstein, A. G. (1979b). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society*, *13*, 187–190.
- Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology*, *105*, 47–59.
- Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *13*, 505–512.
- Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The “general acceptance” of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, *44*, 1089–1098.
- Kassin, S., Tubb, V. A., Hosch, H. M., & Memon, A. (in press). On the “general acceptance” of eyewitness testimony research: A new survey of the experts. *American Psychologist*.
- Lasota, J. A. (1974). *Model rules for law enforcement: Eyewitness identification. Project on law enforcement policy and rule-making*. Washington, DC: Police Foundation.
- Lavrakas, 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. (in press). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit. *Journal of Experimental Psychology: General*.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 732–753.
- Loftus, E. F. (1975). Eyewitness testimony: Does the malleable human memory interfere with legal justice? *Social Action and the Law Newsletter*, *2*, 5ff.
- MacLin, O. H., & Webster, M. A. (in press). The influence of adaptation on the perception of distortions in natural images. *Journal of Electronic Imaging* (Special Section on Human Vision & Electronic Imaging).
- Malpass, R. S. (1990). An excursion into utilitarian analysis. *Behavior Science Research*, *24*, 1–15.
- Malpass, R. S. (1992). “They all look alike to me.” In M. Merrens & G. Brannigan (Eds.), *The undaunted psychologist* (pp. 74–88). New York: McGraw-Hill.
- Malpass, R. S. (1996). Enhancing eyewitness memory. In S. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 177–204). Hillsdale, NJ: Erlbaum.
- Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other races. *Journal of Personality and Social Psychology*, *13*, 333–334.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35.
- Milord, J. (1978). Aesthetics of faces: A (somewhat) phenomenological analysis using multidimensional scaling methods. *Journal of Personality and Social Psychology*, *36*, 205–216.
- Ng, W. J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, *25*, 217–232.
- Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, *11*, 287–307.
- O’Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentine, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine (Ed.), *Cognitive & computational aspects of face processing* (pp. 159–182). London: Routledge.

- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, & Law, 1*, 817–845.
- Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18*, 972–984.
- Rae, G. (1976). Table of A'. *Perceptual and Motor Skills, 42*, 98.
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*, 473–497.
- Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.
- Sergent, J. (1984). Configural processing of faces in the left and the right cerebral hemispheres. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 554–572.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156.
- Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces: A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20*, 125–133.
- Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology, 3*, 137–145.
- State v. Cromedy*, 158 N.J. 112 (1999).
- Studdert-Kennedy, M., Liberman, A. M., Harris, K., & Cooper, F. S. (1970). The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77*, 234–249.
- Swim, J. K., & Shotland, R. L. (1999). *The effect of stereotypes on other-race and same-race identification*. Unpublished manuscript.
- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.
- Thompson, J. (2000). I was certain, but I was wrong. *New York Times, OpEd page*, June 18, 2000.
- United States v. Telfaire*, 152 U.S. App. D.C. 146, 469 F.2d 552, 555 n. 5 (1972).
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161–204.
- Valentine, T., & Bruce, V. (1986). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40*, 300–305.
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology, 44*, 671–703.
- van Wallendaal, L. R., & Kuhn, J. C. (1997). Distinctiveness is in the eye of the beholder: Cross-racial differences in perceptions of faces. *Psychological Reports, 80*, 35–39.
- von Gruenau, M., Dube, S., & Kwas, M. (1996). Two contributions to motion induction: A preattentive effect and facilitation due to attentional capture. *Vision Research, 36*, 2447–2457.
- Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review, 4*, 647–653.
- Wells, G. L., Small, M., Malpass, R. S., Penrod, S., Fulero, S., & Brimacombe, C. A. E.

- (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, 23, 603–647.
- Yasuhara, M., & Kuklinski, T. T. (1979). Category boundary effect for grapheme perception. *Perception & Psychophysics*, 23, 97–104.
- Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, 8, 161–185.

Appendix

Feature and Traits

Features

Nose length: Long–Short
 Nose width: Narrow–Wide
 Mouth width: Narrow–Wide
 Face width: Narrow–Wide
 Lips: Thick–Thin
 Complexion: Light–Dark
 Eyes: Deep–Protruding
 Chin shape: Round–Square
 Ears: Flat–Stick Out
 Eyebrows: Bushy–Thin
 Eye color: Light–Dark

Personality

Practical–Creative
 Cold–Warm
 Weak–Strong
 Dependent–Independent
 Assertive–Submissive
 Relaxed–Tense
 Trusting–Suspicious

Received April 12, 2000

Revision received October 27, 2000

Accepted October 27, 2000 ■