Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 5, 1779–1788

0096-1523/14/$12.00 http://dx.doi.org/10.1037/a0037743

Measuring Sexual Dimorphism With a Race–Gender Face Space

William J. Hopper
University of Massachusetts, Amherst

Kristin M. Finklea and Piotr Winkielman
University of California, San Diego

David E. Huber
University of Massachusetts, Amherst

Faces are complex visual objects, and faces chosen to vary in 1 regard may unintentionally vary in other ways, particularly if the correlation is a property of the population of faces. Here, we present an example of a correlation that arises from differences in the degree of sexual dimorphism. In Experiment 1, paired similarity ratings were collected for a set of 40 real face images chosen to vary in terms of gender and race (Asian vs. White). Multidimensional scaling (MDS) placed these stimuli in a "face space," with different attributes corresponding to different dimensions. Gender was found to vary more for White faces, resulting in a negative or positive correlation between gender and race when only considering male or only considering female faces. This increased sexual dimorphism for White faces may provide an alternative explanation for differences in face processing between White and Asian faces (e.g., the own-race bias, face attractiveness biases, etc.). Studies of face processing that are unconfounded by this difference in the degree of sexual dimorphism require stimuli that are decorrelated in terms of race and gender. Decorrelated faces were created using a morphing technique, spacing the morphs uniformly around a ring in the 2-dimensional (2D) race–gender plane. In Experiment 2, paired similarity ratings confirmed the 2D positions of the morph faces. In Experiment 3, race and gender category judgments varied uniformly for these decorrelated stimuli. Our results and stimuli should prove useful for studying sexual dimorphism and for the study of face processing more generally.

Keywords: face perception, sexual dimorphism, own-race bias, gender perception, face space, multidimensional scaling

Supplemental materials: http://dx.doi.org/10.1037/a0037743.supp

As experts at face processing, we readily extract a large number of socially relevant attributes from faces, despite the high degree of visual similarity between different faces (Baudouin & Tiberghien, 2002; Brown & Perrett, 1993; Bruce, Ellis, Gibling, & Young, 1987; Bruce & Langton, 1994; Burton, Bruce, & Dench, 1993; Ekman & Oster, 1979; Etcoff, 1986). Many studies investigate variation along just one of these attributes, assuming that the chosen faces are approximately equal in other attributes. However, there may be unintended correlations between the attribute of interest and other attributes of the face stimuli, particularly if such correlations exist in the population. Without understanding the nature of the correlations between attributes, developing accurate theoretical accounts of any behavioral phenomenon will be difficult, as attributes other than those specifically isolated in studies of face processing may provide an alternative explanation.

This article was published Online First August 25, 2014.

William J. Hopper, Department of Psychology, University of Massachusetts, Amherst; Kristin M. Finklea and Piotr Winkielman, Department of Psychology, University of California, San Diego; David E. Huber, Department of Psychological and Brain Sciences, University of Massachusetts, Amherst.

Correspondence concerning this article should be addressed to William J. Hopper, Department of Psychology, Tobin Hall, 135 Hicks Way, University of Massachusetts, Amherst, Amherst, MA 01003. E-mail: whopper@psych.umass.edu

Here, we present an example of such a correlation that arises from differences in the degree of sexual dimorphism for faces of different races. To minimize demand characteristics by avoiding explicit reference to race or gender, we collected similarity ratings for pairs of faces. Sexual dimorphism was measured using these similarity ratings to create a multidimensional "face space" in which each attribute corresponds to a dimension and each faces position along a dimension characterizes the degree to which that face expresses that attribute (Johnston, Milne, Williams, & Hosie, 1997; Rhodes, Byatt, Tremewan, & Kennedy, 1997; Valentine & Endo, 1992; Valentine, 1991).

There are a wide variety of face spaces, although they can be categorized as ones that are theoretically derived versus descriptive. A theoretically derived face space specifies, in advance, the physical properties underlying each dimension (e.g., distance between the eyes), and based on these assumptions, any face can be located in the face space. In contrast, statistical techniques, such as multidimensional scaling (MDS; Kruskal & Wish, 1978), describe the locations of a specific set of normed faces without making any assumptions about the psychological or physical properties that underlie each dimension. The current study sought to measure differences in the degree of sexual dimorphism for difference races, and so a descriptive approach was used.

Intuitively, the dimensions of race and gender are perhaps the most informative dimensions along which faces vary. For this reason, face spaces often capture one or the other of these two dimensions (Busey, 1998; Byatt & Rhodes, 2004; Johnston et al., 1997; Valentine & Endo, 1992). To the best of our knowledge, a face space has never been created that includes both of these dimensions. This is an important oversight because studies that used explicit gender judgments find differences between races. For instance, male faces of some races are perceived as more masculine than the male faces of other races (Galinsky, Hall, & Cuddy, 2013; Goff, Thomas, & Jackson, 2008; Johnson, Freeman, & Pauker, 2012). This correlation between race and masculinity gives rise to the possibility that reported effects of one of these dimensions might reflect unintended variation of the other dimension. More generally, these effects suggest that faces of different races may differ in the degree of sexual dimorphism. However, a study of this this hypothesis based on explicit gender ratings is problematic—participants may attempt to be unbiased in their treatment of each race, or alternatively, racial stereotypes may artificially produce differences. The current study avoided these issues by collecting simple similarity ratings between pairs of faces, which places an emphasis on perceptual differences rather than race or gender differences.

We started with a set of 40 faces that was evenly split across the two genders and across two races (i.e., 10 faces for each combination of race and gender). We did not attempt to identify the physical characteristics that capture these dimensions. Instead, our goal was to locate each face within the psychological dimensions of race and gender through the use of similarity ratings. By locating a set of faces in this psychological two-dimensional (2D) space, we provide normed stimuli that can be used for categorization studies (Nosofsky, 1986) or any other study of the role of race or gender in face processing. As reported here, we normed these faces with MDS as applied to similarity judgments for pairs of faces. To avoid contamination of the results owing to the "own-race bias" (e.g., Meissner & Brigham, 2001), we used participants that were highly experienced with both races. Otherwise, the race dimension might reflect perception of one's own race at one extreme versus "other-race" at the other extreme (i.e., the absence of the characteristic that defines your own race). To this end, we chose White and Asian as our two races, considering that the undergraduate population at the University of California, San Diego is evenly split between these two races.

Experiment 1

Method

Participants. All participants ($N = 162$) were undergraduate students at the University of California, San Diego, and participated as partial fulfillment of course credit. The sample of participants was 33% male, 33% White, and 67% Asian. Twenty-eight of the participants viewed all 780 possible pairs of faces. The remaining participants viewed only 390 pairs of faces, such that two participants (independently) viewed the two halves from each random order of face pairs.

Materials. The stimuli consisted of 40 pictures (300 × 400 pixels, grayscale) representing all four race–gender categories: 10 White men, 10 White women, 10 Asian men, and 10 Asian women. The original photographs were provided by Dr. Kang Lee[1] and used previously by Ge et al. (2009). The black-and-white photographs, similar to passport pictures, were chosen because they were well-centered, neutral-expression headshots without facial hair, eyeglasses, or jewelry, taken under similar lighting and background. Visually controlled stimuli such as these are important when using MDS, as nuisance dimensions such as brightness or viewing angle can emerge as primary dimensions of the solution. There are many physical attributes that may contribute to the perception of race and gender, although some, such as hair style, may reflect cultural differences rather than intrinsic attributes of faces. To eliminate the influence of hair style, we altered the photos by using an oval filter for each image to delineate the interior versus the exterior. One face was chosen from each of the four race–gender categories, and the exteriors from these four faces were averaged. The interior of all 40 faces was then placed inside this average exterior (see Figure 2 for examples of the stimuli from each of the four categories of race–gender). This procedure eliminated the possibility that subjects relied on features external to the face when judging similarity, including head shape to some degree, but maintained a realistic face that can be processed in the same way as unaltered faces.

Procedure. Based on these 40 face images, there were 780 unique face pairs. On each trial, a unique face pair was presented to participants with the faces located beside each other, with position (left or right) randomly determined on each trial. Participant responses were collected on the keyboard using the numeric keys 1–7. Stimuli presentation as well as participant response data were controlled and collected by E-Prime (Psychology Software Tools, Inc., Sharpsburg, PA). Participants rated the similarity of face pairs on a 7-point scale (1 = *very different*, 7 = *very similar*). Each face pair remained on the screen for 3 s before similarity ratings were given. Participants received a short break after every 50 pairs of faces in order to minimize fatigue.

Results

Similarity rating data were collapsed across participants who viewed the entire set of 780 face pairs and participants who viewed half of the face pairs. An MDS solution was calculated using nonmetric MDS and the normalized stress criterion (Kruskal & Wish, 1978; Takane, Young, & De Leeuw, 1977). Although race and gender captured most of the variance, a six-dimensional solution was chosen to factor out face attributes other than race and gender. The resulting six-dimensional MDS solution had an r^2 value of .946 and a stress value of .070, indicating of a great degree of correspondence between ideal and actual distances (Sturrock & Rocha, 2000).

As noted by Busey (1998), one test for the quality of an MDS solution is whether the resulting dimensions are interpretable. As reported elsewhere (Finklea, 2008), five of the six dimensions were statistically identifiable[2]; however, the focus in the current study is placed on the race and gender dimensions of the solution, which were visually identifiable as the first and second dimensions

[1] Dr. Kang Lee, Professor and Director, Institute of Child Study, University of Toronto, 45 Walmer Road, Toronto, Ontario, Canada M54 2X2. E-mail: kang.lee@utoronto.ca

[2] Experiment 1 is adapted from Experiment 3 of the unpublished doctoral dissertation of Kristin M. Finklea (2008), which compared Dimensions 3 through 6 to a variety of physical measurements of the faces. For instance, Dimension 3 correlated with eyebrow darkness.

of the MDS solution, respectively (see Figure 1). All White faces were located on the negative side of the *X* axis, while all Asian faces were located on the positive side of the *X* axis, suggesting that the first dimension of the solution represents race. Similarly, male and female faces can be perfectly separated by a value on the *Y* dimension (e.g., the horizontal line $Y = .1$), suggesting that the second dimension of the solution represents gender. However, the Asian faces were closer to the neutral gender line, whereas the White faces varied more greatly in terms of gender. This resulted in a negative relationship between race and gender for the male faces but a positive relationship for the female faces: The mean position on the gender axis was significantly higher for the White male faces than the Asian male faces, $t_{18} = 4.18$, $p < .001$, but the opposite was true for the female faces, $t_{18} = 1.95$, $p < .05$.

Our aim was to measure differences in sexual dimorphism in an unbiased manner. However, there were twice as many Asian participants as White participants. Thus, if the two races perceived these faces differently, our results would be biased toward an Asian participant interpretation of these stimuli. To check whether the two races perceived these faces differently, we compared the similarity ratings given by Asian and White participants. As compared to Asian participants, White participants rated pairs of Asian faces to be more similar, $t_{161} = 24.66$, $p < .001$. However, White participants also rated White faces to be more similar, $t_{161} = 9.61$, $p < .001$. In other words, it appears that White participants used the measurement scale differently by indicating greater similarity in general. Aside from this main effect of similarity, we did not find any participant-race effects, and separate 2D MDS solutions for Asian and White participants were nearly identical. Dimension 1 from the Asian MDS solution highly correlated with Dimension 1 from the White MDS solution, $r_{38} = .996$, $p < .01$, and Dimension 2 from the Asian solution highly correlated with Dimension 2 from the White Solution, $r_{38} = .983$, $p < .01$.

Discussion

In this experiment, 40 faces were described with a race and gender face-space using MDS applied to paired comparison similarity ratings. Along the gender dimension, there was greater variability for the White faces (i.e., greater sexual dimorphism), and this resulted in a correlation between race and gender when considering only male or only female faces. This study adds to a growing body of literature showing perceptual interactions between the race and gender of faces. Although 40 is a relatively small sample of faces, similar results have been reported by others (Galinsky et al., 2013; Goff et al., 2008; Johnson et al., 2012). We have extended these finding by using similarity ratings and MDS to characterize race and gender relationships between Asian and Whites without invoking demand characteristics that might arise from explicit reference to race or gender. The demographics of the university students sampled in this experiment were well-balanced between these two races. Because these students have considerable experience with both races, the separate MDS solutions for participants of each race were nearly identical (i.e., we did not see any own-race bias effects in these data). Thus, our results provide a normative description of the difference in the degree of sexual dimorphism.

Also using MDS, the same conclusion of greater variability for White faces was found by Byatt and Rhodes (2004). However, that study only used male faces and all of the participants were White, leading to an interpretation that the greater variability for White faces was the result of an own-race bias to be more sensitive to differences among one's own race. Because our study used faces of both genders, and participants of both races, we were able to determine that the greater variability for the White faces was in terms of gender perception, rather than an artifact of own-race bias. More specifically, because we utilized participants of both

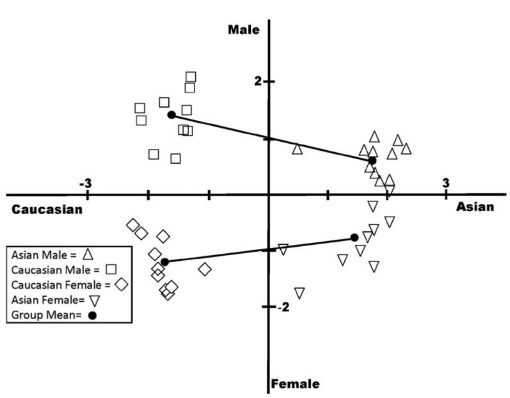

Figure 1. Scatterplot of the multidimensional scaling solution from Experiment 1. The filled black circles show the averages for each of the four combinations of race and gender. The lines connecting these averages highlight the correlations between race and gender when only considering faces of one gender.

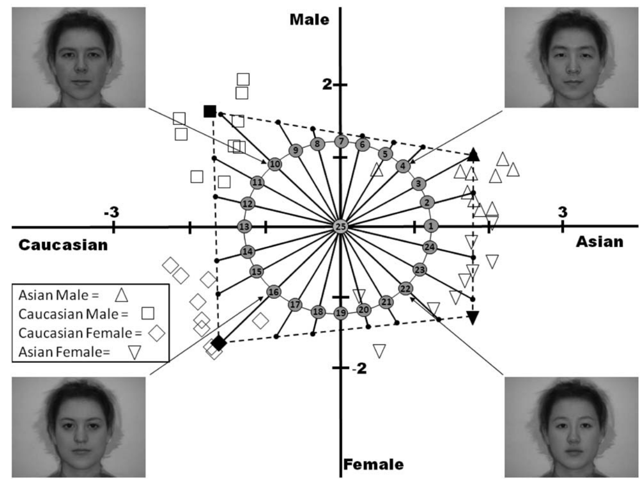

Figure 2. Graphical representation of stimuli creation process designed to decorrelate race and gender based on the multidimensional scaling space shown in Figure 1. Using the real faces from Experiment 1, prototypical faces were selected from each race–gender combination, as indicated by the filled symbols. The prototypes were connected with morph sequences to create a quadrilateral boundary. Next, 12 spoke morph sequences were created connecting morphs on opposite sides of the quadrilateral, and three morphs were created from each spoke to produce morphs on opposite sides of a race–gender ring and a best estimate for the neutral origin face. The 12 estimates of the origin face were then averaged, and the 24 ring faces (shown by the numbered gray circles) plus the origin face were used in Experiment 2. The four face images show the morph faces from the race–gender diagonals (morph numbers 4, 10, 16, and 22).

races, we demonstrated that greater gender variability among White faces was perceived by participants of both races, not only Whites.

As noted in the introduction, understanding the natural correlations between face attributes is important to developing theories of face processing, as they may provide alternative explanations or suggest reinterpretations of existing theoretical accounts. As one example, the finding that Whites are perceived to have greater sexual dimorphism suggests an intriguing new interpretation of the inappropriate cue utilization explanation of own-race bias (Shepherd & Deregowski, 1981). Cue utilization holds that some facial cues (e.g., eye color) may be more useful for distinguishing between individuals of some races as compared to other races. By having more experience with one's own race, an observer learns to pay attention to the cues most useful for their own race and would likely use the same cues to discriminate faces of another race. However, if these cues are not as useful for faces of other races, discriminability and recognition may suffer. A cue is more useful if there is greater variability between individuals for that cue, and our results suggest that gender (e.g., degree of masculinity or femininity) may be a race-specific cue. Thus, the own-race bias effect of greater recognition and discrimination accuracy of White faces by White observers with little experience with Asian faces may reflect greater attention to gender cues, which are less useful for Asian faces. Conversely, Asian observers with little experience with White faces may have learned to pay more attention to cues other than gender, to the detriment of discrimination between White faces.

The hypothesis that gender may be a race-specific cue implies that gender judgment accuracy should be higher for Whites faces, regardless of the race of the observer. Indeed, participants of both races (White and Asian) are quicker and more accurate when categorizing the gender of White faces than Asian faces (Zhao & Bentin, 2008; but see also O'Toole, Peterson, & Deffenbacher, 1996). Importantly, this hypothesis is not an all-or-nothing use of gender cues—in all cultures, it is important to identify whether someone is male or female. Instead, the difference occurs in the extent to which gender cues are used to distinguish between two faces of the same sex. Thus, even though gender cues are used for gender categorization for faces of all races, our results suggest that gender cues will be more useful for individuation of White faces than Asian faces.

Our results may also lead to new interpretations of studies that used participants of only one race. In such situations, it may be that differences in gender variability between the races underlie any observed own-race bias effects. For instance, using participants from just one race, Lovén et al. (2012) found an own-race advantage only for recognition of female faces, which suggests that gender cues played an important role. Similar to our conclusions, the authors concluded that face gender modulates visual attention and memory. However, our results take things a step farther, suggesting that some own-race bias effects may reflect gender-variability differences in the face stimuli of each race. For instance, perhaps the female faces in the Lovén et al. study varied along the gender dimension more for the own race than the other race, and this was the cause of the own-race bias effect. As the authors suggest, more information may have been extracted from female own-race faces than female other-races faces, and our findings suggest that this information may be gender information itself.

It is evident that experiments investigating the nature of race should carefully consider whether the chosen stimuli coincidentally vary along other important dimensions such as gender. Paradoxically, these correlations were made apparent when considering only male faces or only female faces. Thus, attempts to control the role of gender by selecting faces of only one gender may, in truth, create a correlation between race and gender. Nevertheless, an important conclusion from Experiment 1 is not this specific correlation, but rather that such correlations can exist in general when using a random sample of faces selected to vary along one dimension of interest (e.g., race).

As discussed earlier, this difference in the degree of sexual dimorphism may pose a problem for many studies that only consider race or gender. To unconfound investigations of race or gender from the interaction between race and gender requires stimuli that are decorrelated along these dimensions. Using morphing software, we developed decorrelated stimuli by creating faces that were evenly spaced around the 2D race–gender plane. The resultant ring of faces had no preferred axis of variation, and the correlation between race and gender was eliminated for these morph faces. Such decorrelated faces should be useful for examining the consequences of sexual dimorphism (e.g., the hypothesis that gender is a race-specific cue) and, more generally, for well-controlled studies of face processing.

Experiment 2

We created a ring of faces in the 2D race–gender plane, thus eliminating any correlation between race and gender. As in Experiment 1, the locations of these faces were determined through paired comparison similarity ratings and MDS.

A race–gender ring of faces was created using morphing software. Four of the real faces from Experiment 1 were selected as being prototypical based on their MDS race–gender locations (see Figure 2). Morph sequences between these prototypes created morph faces around a quadrilateral. New morph sequences were created that connected morph faces on opposite sides of the quadrilateral to create a ring of 24 equally spaced race–gender morph faces. Because a circular ring has mirror symmetry in all directions, we anticipated that the 2D MDS solution for these morph faces would be an arbitrary rotation/reflection of the MDS space found in Experiment 1. Thus, the main goal of the MDS solution for these 24 morph faces was to verify their positions relative to each other. The absolute race–gender positions for these morph faces were then determined by rotating/reflecting the Experiment 2 MDS solution to align with the Experiment 1 MDS solution.

Method

Participants. Sixty-seven participants were recruited from the University of California, San Diego, undergraduate subject pool. Subjects were given one unit of credit that could be applied either toward class participation requirements or extra credit opportunities in undergraduate psychology classes.

Materials. Twenty-five morph faces (see Supplemental Materials) were created using four of the real faces from Experiment 1 that were prototypical of the four race–gender combinations:

Asian male, Asian female, White male, and White female. Based on the Experiment 1 MDS solution, these prototypical parent faces were selected for the property of lying near one of the four race–gender diagonals while being as far from the origin as possible. This served to maximize the radius of the circle contained within the box defined by these four faces. There was some degree of flexibility in choosing these faces, and we also made sure that the selected faces were not otherwise unusual or overly memorable.

Using the FantaMorph (Abrosoft, Haidian, Beijing, China) software package, morph sequences between prototype faces created a quadrilateral of morph faces, with each side of the quadrilateral representing combinations of two of the prototype faces. FantaMorph creates transitions between two static images using a method known as "field morphing" to linearly interpolating the pixels near specifically selected points on each image (Beier & Neely, 1992; Steyvers, 1999). With faces, these points are selected to outline landmark features such as the shape of the nose, mouth, eyes, eyebrows, and iris. Points are placed on each face in a one-to-one correspondence, until the entire face is outlined. This method is commonly used in studies to make combinations of faces with varying amounts of information from each parent face, creating a "path" of intermediary faces through the face space, between the two parent faces (Fox & Barton, 2007; MacLin, Peterson, Hashman, & Flach, 2009).

Using trigonometric calculations applied to the Experiment 1 MDS solution, 24 morph faces around the quadrilateral were identified. These morph faces were selected to lie on 12 "spokes" in the 2D race–gender plane spaced 15° apart and passing through the origin. Each spoke connected morph faces on opposite sides of the quadrilateral, and each of the morph sequences along a spoke was used to create two new morphs both located 1.35 units in MDS space from the origin, with this radius ensuring that the ring was entirely contained within the quadrilateral. Rather than some other number of ring faces, 24 was chosen because it is divisible in several different ways, which makes these stimuli useful for a variety of different experimental designs. Finally, each of the 12 spoke morph sequences was used to create an estimated origin face, and these 12 estimates were averaged together to produce a single best estimate of the origin face.

Procedure. The procedure was identical to Experiment 1 except as noted. With 25 rather than 40 faces, only 300 trials were necessary to collect similarity ratings for all pairs of the morph faces. All images had grayscale coloring, a width of 8.67 in. and a height of 6.67 in., and were displayed at a 620-pixel × 480-pixel resolution.

Results

The same MDS algorithm used in Experiment 1 was applied to the similarity ratings, although the solution was limited to two dimensions rather than six, considering that the 25 morph faces were created from combinations of four race–gender prototype faces. This MDS solution had an r^2 value of .9603 and a stress value of .0798.

Remarkably, the MDS locations of the 24 ring faces were approximately in a ring (see Figure 3), and the ordering of these faces was exactly as predicted from the process used to create them. In Experiment 1, race accounted for more variance than

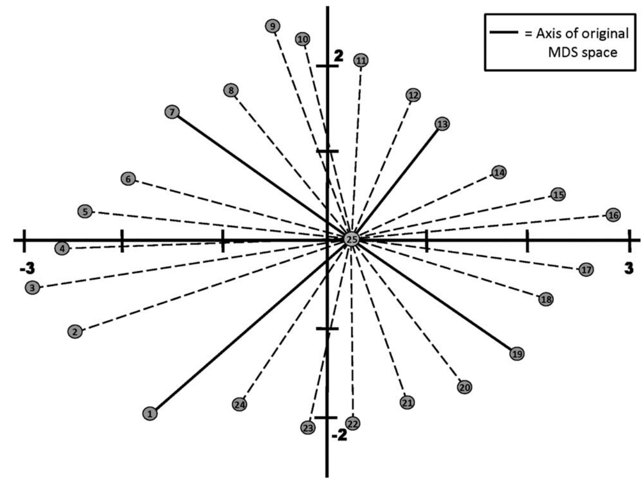

Figure 3. Results of the multidimensional scaling (MDS) solution based on similarity ratings between the 25 newly created morph faces. Solid radial spokes connecting the origin face (25) to faces 1, 13, 7, and 19 represent boundary lines for gender and race, respectively. Demonstrating this race and gender were successfully decorrelated for this ring of morph faces, this MDS solution is an arbitrary reflection/rotation of the Experiment 1 MDS solution.

gender for the 40 real faces, whereas these morphed faces were designed to be equidistant in the race–gender plane (i.e., race and gender should be equally important). Thus, as expected, the 2D MDS solution for Experiment 2 was an arbitrary rotation/reflection of the race–gender plane. To align this solution with the Experiment 1 solution, the positions of the 24 ring faces were linearly transformed by (a) reversing the first dimension; (b) subtracting the average of the 24 locations (i.e., centering); (c) rotating the angle of all faces by the average difference between the actual angles and the desired angles used to create the faces; and (4) rescaling all locations so that the average distance to the origin was 1.0. These transformed locations are reported both in Cartesian and polar coordinates in Table 1.

For these transformed locations of the 24 ring faces, the 99% confidence interval for the mean radius ranged from 1.08055 to 0.919454, assuming a t distribution with 23 degrees of freedom. Ideally, the faces should be exactly 15°, apart and the results did not greatly vary from this ideal. The 99% confidence interval for the mean angle between each face (with vertex at the origin), assuming a t distribution with 23 degrees of freedom, ranged from 12.06° to 17.93°.

These results rely on transformations of the MDS solution (which is, itself, a transformation of the raw data). Alternatively, equal spacing for the 24 ring faces can be assessed using the raw similarity ratings for pairs of adjacent faces. The 99% confidence interval for the mean similarity between adjacent faces ranged from 5.46 to 6.03, assuming a t distribution with 23 degrees of freedom. A repeated-measures one-way ANOVA indicated that some adjacent faces were rated more similar than others ($F_{23,\ 1472} = 2.397$, $p < .01$). However, with 24 levels to the independent variable, a significant omnibus test is unsurprising. Suggesting a great deal of uniformity to the similarity ratings of adjacent faces, only one of the 276 different post hoc comparisons between different pairs of adjacent faces was significant.

Table 1
Cartesian and Polar Coordinates of the Ring Faces Created in Experiment 2

Face	*X* (race)	*Y* (gender)	θ	*r*
1	1.158	0.004	0.003	1.158
2	1.060	0.515	0.452	1.179
3	1.015	0.808	0.673	1.297
4	0.781	0.836	0.820	1.144
5	0.576	0.887	0.995	1.058
6	0.329	0.848	1.200	0.909
7	−0.051	0.923	1.626	0.924
8	−0.296	0.801	1.925	0.854
9	−0.659	0.881	2.213	1.100
10	−0.699	0.735	2.331	1.014
11	−0.777	0.475	2.592	0.911
12	−0.805	0.195	2.904	0.828
13	−0.766	−0.002	−3.139	0.766
14	−0.752	−0.343	−2.714	0.827
15	−0.836	−0.611	−2.511	1.035
16	−0.912	−0.856	−2.388	1.251
17	−0.632	−0.949	−2.158	1.140
18	−0.399	−0.914	−1.982	0.997
19	−0.116	−0.994	−1.687	1.000
20	0.156	−0.932	−1.405	0.945
21	0.378	−0.802	−1.130	0.887
22	0.617	−0.693	−0.844	0.928
23	0.760	−0.558	−0.633	0.943
24	0.867	−0.255	−0.286	0.904

Discussion

The MDS solution for these morph faces validated the procedure used to create them. The correspondence between the predicted spatial relationship of faces created by morphing manipulations and the relationships observed in new MDS solution based on similarity ratings of the morphed faces show that the morphing procedure based solely on MDS position produced a directly corresponding change in the psychological similarity between the faces. As seen in Figure 3, there was no longer any correlation between race and gender for these morph faces. This decorrelation of race and gender was achieved by designing the faces to lie on the circumference of a race–gender circle, which resulted in an arbitrary reflection/rotation of the Experiment 1 race–gender 2D plane (e.g., no longer was race more important than gender). Furthermore, the order of the faces was as predicted and the faces were approximately uniform in their spacing. Most importantly, the gender variation among White faces was not significantly greater than among Asian faces (i.e., for these stimuli, differences in sexual dimorphism for faces of different races was eliminated). However, the similarity ratings collected in both Experiment 1 and Experiment 2 were made without any explicit reference to race or gender, and the dimensions of the solution need to be empirically validated. To address this need, race and gender categorization judgments for each of the 24 ring faces were collected in Experiment 3 and were compared to predictions derived from each faces position in MDS space.

Experiment 3

The first two dimensions of the MDS solution in Experiment 1 were assumed to capture race and gender, considering that the experiment used faces sampled from each of the four race–gender categories and that each race and gender combination was located in a distinct quadrant of the solution. However, it is conceivable that other face attributes perfectly covaried with race and gender for this sample of faces. If so, the morph faces created in Experiment 2 may have varied along dimensions other than race and gender. We tested this possibility by asking participants to give explicit race/gender categorization judgments for the 24 ring faces. This allowed us to ascertain whether these morph faces truly were evenly spaced along the dimensions of race and gender.

This experiment not only verified the nature of the dimensions, but also served to test whether "categorical perception" for race and gender was a key component of the similarity judgments for these faces. Categorical perception results in greater sensitivity to small perceptual differences for stimuli that lie on either side of a categorical boundary and insensitivity to perceptual differences for stimuli that are from the same category. This can be verified either with category judgments (e.g., a nonlinear step function for category judgments as a function of continuous perceptual change between categories) or with same/different judgments (e.g., greater accuracy on the boundary for same/different judgments as a function of continuous perceptual change between categories). Evidence of categorical perception has been observed both with gender (Bülthoff & Newell, 2005; Campanella, Chrysochoos, & Bruyer, 2001) and race (Levin & Angelone, 2002) by using morphing to continuously change between categories. However, in the absence of norming data for the morphs, the mapping between equal morph steps and psychological representations is unknowable. The stimuli developed in Experiment 2 provide a unique opportunity to examine categorical perception because the psychological locations of these stimuli were determined with similarity judgments and MDS. If categorical judgments of race and gender were an integral component of the similarity judgments in Experiment 2, then a plot of the observed race and category judgments as a function of the *X* and *Y* values from Table 1 should be roughly linear (noncategorical) rather than a steep threshold function. In other words, if two faces just on either side of the race boundary were judged to be dissimilar by virtue of coming from different race categories, there would be a stretching of the similarity-based *X* dimension near the boundary that counteracts categorical perception.

Aside from predicting the absence of categorical perception for these evenly spaced morph faces, we also expected Experiment 3 to deviate from the literature (Zhao & Bentin, 2008), by failing to produce a gender categorization advantage for White faces as compared to Asian faces. We hypothesize that Zhao and Bentin observed such an advantage because of greater gender variability for White faces (e.g., the average male White and average female White are farther from the neutral gender categorization boundary). However, because these morph faces are evenly spaced in terms of gender, regardless of race, we expected that gender categorization responses (both choice and reaction time [RT]) would not be influenced by the race of the face. Thus, gender categorization responses should only depend on the *Y* values, but not the *X* values.

Method

Participants. Forty-nine participants were recruited from the University of California, San Diego, undergraduate subject pool. Subjects were given one unit of credit that could be applied either toward class participation requirements or extra credit opportunities in undergraduate psychology classes.

Materials. The stimulus materials were the same 24 ring faces used in Experiment 2.

Procedure. Participants were instructed to make decisions as quickly as possible without sacrificing accuracy. Two blocks of faces were presented, each containing all 24 morphed faces, presented in random order. For one block, participants were asked to classify each face as male or female, and in the other block, they were asked to classify each face as Asian or White. The order of the blocks was selected randomly for each participant. Each face was presented for 2,000 ms, and if participants failed to make a classification judgment by the end of the presentation duration, they were prompted with a question mark. The next trial did not commence until an answer was provided.

Results

RT differences between faces were analyzed to test whether participants slowed their categorization responses for faces near category boundaries. There were no officially correct responses for this categorization task, and thus, RT analyses began by including all responses. Means and standard deviations of RT were found for each participant in each type of categorization task. Trimming was necessary because, occasionally, subjects spent a very long time (e.g., 20 s) giving their responses. Responses with a RT greater than 4 standard deviations about the mean were discarded from the RT analysis and from subsequent analysis ($n = 6$). Because only one data point was collected per subject per each face in each categorization task, we had to choose whether to perform trimming on the basis of faces or on the basis of subject. Considering that subject differences were larger than face differences, these means and standard deviations were calculated for each subject, collapsing across faces. Because this trimming resulted in missing cells for a within-subjects analysis, a between-subjects analysis was the only viable manner for assessing whether RT differed between faces. No significant difference in RT was observed between faces in the gender categorization task, $F_{23,1149} = 1.34$, $p = .13$, or in the race categorization task, $F_{23,1149} = 1.4$, $p = .09$.

As seen in Figure 4, the *X* and *Y* values in Table 1 provided a remarkably accurate prediction for the race and gender categorization judgments. This is notable in that the *X* and *Y* values were based on similarity judgments without any explicit reference to race or gender. In generating these predictions, the $X = 0$ and $Y = 0$ lines were assumed to correspond to 50/50 race and gender categorization responses, respectively. These predicted categorical boundaries were very close the observed data. In truth, the $X = 0$ and $Y = 0$ lines are somewhat arbitrary because they were essentially defined by the mean of the Experiment 1 stimuli along each dimension. That these default boundaries predicted the results of Experiment 3 suggests that the stimuli used in Experiment 1 were on-average of neutral gender and neutral race. Aside from the locations of the category boundaries, the *X* and *Y* values also predicted the form of the transitions between races and gender—with increasing values of *X* or increasing values of *Y*, categorization judgments smoothly transitioned from proportions near zero to proportions near one.

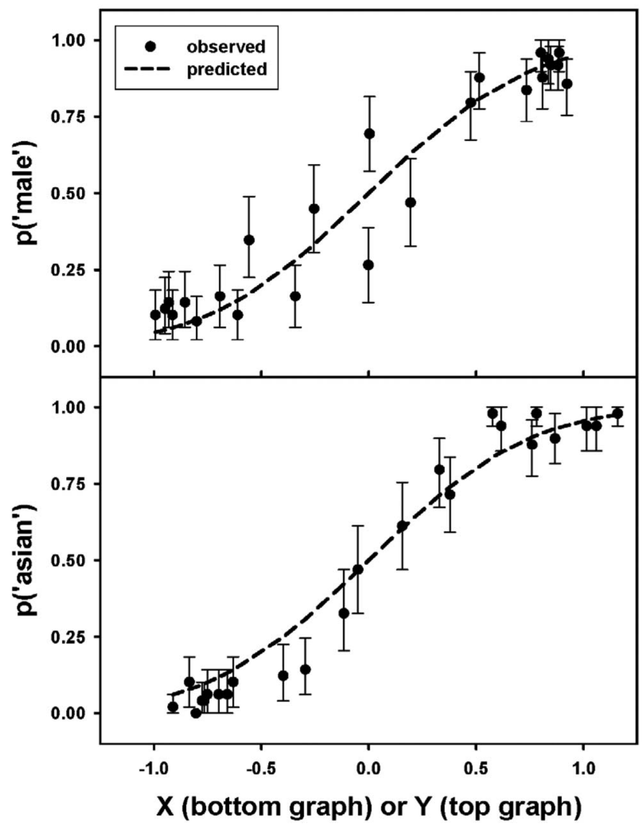

Figure 4. Categorization proportions from Experiment 3, plotted against the *X* (race) and *Y* (gender) values calculated from Experiment 2 (see Table 1). Error bars represent a 95% binomial confidence interval. The dashed line shows predicted categorization frequencies based on the *X* and *Y* values, assuming that the boundary lines occur at $X = 0$ and $Y = 0$.

The predicted lines in the figure are not entirely a priori predictions because one parameter was adjusted to fit the data. This parameter captured noisy perceptions/decisions by adding a standard normal deviation to the difference between the perceived *X* or *Y* position and the zero line demarking the category boundary. Thus, the model estimates the proportion of times that perception of the face is stochastically sampled, according to a normal distribution, to lie on one side of the boundary versus the other. Because the same standard deviation was used for both judgments, the predicted lines for both graphs are identical and are specified by a zero mean normal distribution with a free parameter for the standard deviation. This standard deviation parameter determines the degree of categorical perception—if the standard deviation is small, the function will be step-like, whereas a large standard deviation corresponds to a more linear function. The likelihood of the data was maximized using the binomial likelihood ratio test, which provides a chi-square goodness-of-fit statistic. Chi-square was minimized at 121.9, with a best-fitting standard deviation (sigma) of .5885. This single parameter accounted for over 95% of the variance ($r^2 = .951$) across the 48 conditions. Considering that nearly all of the *X* and *Y* values fall within a distance of one from the zero category boundary lines, this indicates that all of the faces were within two standard deviations of the category boundary, corresponding to relatively noncategorical perception (a line could also produce a reasonable account of these data).

Discussion

Experiment 3 confirmed that the dimensions of the 2D face space in Experiment 2 were, in fact, race and gender—the *X* and *Y* values for the morph faces produced a fairly accurate prediction of race and gender category judgments, without an interaction between race and gender such as found by Zhao and Bentin (2008).[3] When plotted as a function of the *X* and *Y* MDS values, these category judgments were relatively noncategorical (i.e., more linear). This suggest that the similarity ratings collected in Experiments 1 and 2 were sensitive to categorical perception of race and gender—if participants relied on knowledge about race and gender categories to perform their similarity ratings, then increased perceptual dissimilarity between faces that span the category boundary (relative to faces that are the same distance apart, but are in the same category) would be incorporated into the MDS solution. Thus, when classifying the faces, no sharp transition in categorization frequency is seen near the category boundary, supporting the assertion that these faces were evenly spaced in the 2D face space. This can be contrasted with experiments reporting race/gender categorical perception when plotting the results as a function of morph steps rather than steps of equal similarity (Bülthoff & Newell, 2005; Campanella et al., 2001; Levin & Angelone, 2002). Our results do not conflict with these reports of categorical perception. Instead, the point is that categorical perception occurs when the results are analyzed as a function of equal increments of physical change, but categorical perception is eliminated when the results are analyzed as a function of equal increments of psychological similarity.

General Discussion

Psychological experiments into the nature of face processing often use a selection of faces chosen to vary along a specific attribute. For instance, in studying own-race bias, an experiment might compare responses to a sample of Asian faces versus a sample of White faces. Similarly, many experiments on attractiveness compare Asian and White faces. However, face attributes such as race may be correlated with other important attributes such as gender. The results of Experiment 1 provide an example of such a correlation using an equal number of real faces from the four combinations of Asian/White and male/female. Critically, our participants had considerable experience with both races and there were no own-race bias effects in our data. Based on an MDS solution for similarity ratings between all pairs of faces, the first dimension of variation captured race and the second captured gender. However, the faces were not evenly distributed in each dimension of the face space, with White faces varying more greatly in terms of gender.[4] In other words, we observed greater sexual dimorphism for White faces as compared to Asian faces. This was determined from perceptual similarity ratings without explicit reference to race or gender, thus minimizing the role of demand characteristics. As a result of this greater sexual dimorphism for White faces, there were reliable correlations between race and gender for these faces when considering only male faces or only female faces.

Additional studies of this sexual dimorphism difference, and its consequences, will require face stimuli that are evenly spaced along the psychological dimensions of race and gender. Such stimuli were created in Experiment 2. The MDS solution from Experiment 1 was used to identify four prototype faces and these faces were combined to different degrees using morphing software to create a set of morph faces that were uniformly spaced around a race–gender ring.[5] In contrast to the unevenly distributed real faces, the MDS solution for this ring of morph faces had no preferred orientation (i.e., the locations of the morph faces in the 2D face space were uncorrelated). In short, although the real faces exhibited sexual dimorphism differences between the races, this effect was eliminated for the morph faces.

The behavioral responses for Experiments 1 and 2 were similarity ratings, without explicit reference to race or gender. To confirm that the MDS dimensions corresponded to race and gender, race and gender category judgments for the ring of morph faces were collected in Experiment 3. The boundary lines between races and between genders occurred at the values predicted by MDS. Demonstrating that the morph faces were evenly spaced in terms of explicit gender and race ratings, the category judgments were relatively smooth and linear (noncategorical) when plotted as a function of the MDS values.

The finding of greater sexual dimorphism for White faces may give rise to many novel hypotheses regarding face processing phenomena that were previously assumed to reflect the singular dimensions of race or gender. For instance, we developed the novel hypothesis that own-race bias might, in part, reflect the tendency to rely more heavily gender cues if one's prior experience is with a race that exhibits greater sexual dimorphism. Simply put, if gender differs more greatly between individuals, then gender will be a more effective cue for differentiating between individuals, and general use of this strategy will produce a relative deficit when discriminating faces from races where the magnitude of sexual dimorphism is

[3] The model's fit to the gender categorization data was not as good as for the race categorization. This may be an artifact of the assumption made by MDS that dimensions are orthogonal. If the psychological factor underlying the second dimension (gender) was not in truth orthogonal to the psychological factor underlying the first dimension (race), this will slightly tilt the MDS determined values of the second dimension in relation to the true values of gender, such as revealed by using categorical gender judgments.

[4] Because the faces used a common exterior, peripheral race/gender cues, such as hairstyle and brow/chin shape, were not identifiable. As such, our results primarily relate to frontal views in which the observer makes judgments based on interior (i.e., configural) face attributes. However, these are the aspects of face processing that are commonly studied in the literature.

[5] Blends of the four prototype faces were created using software constrained to combine two faces at a time. These four-way blends were achieved by morphing together two prototype faces that were then combined with blends of the other two prototype faces. This is more easily achieved with software that directly calculates blends of four or more faces. An advantage of directly morphing multiple faces is that the creation of equally spaced faces could be extended to situations involving more than two dimensions.

smaller. The MDS solution and morphed stimuli from Experiment 2 could be used to test this hypothesis by comparing discrimination performance for participants of different races when viewing faces that varied in gender but not on race, and when viewing faces that varied on race but not gender. This gender-cue-utilization hypothesis predicts that White participants will have higher discrimination performance for the faces that vary solely in terms of gender, and that own-race bias effects for both Asian and White viewers should be minimized or eliminated when distinguishing between the faces that vary solely in terms of race. Beyond own-race bias, our results might give rise to novel hypotheses of face attractiveness. For instance, it has been observed that Asian male faces are rated as less attractive than White male faces and this has been attributed to the racial stereotype of low-dominance Asians (Galinsky et al., 2013). Our results suggest an alternative hypothesis of greater femininity for Asian male faces and our decorrelated stimuli could be used to differentiate between these competing hypotheses.

Aside from our conclusions regarding sexual dimorphism, the broader message is that stimuli need to be controlled in terms of perceptual similarity in general, rather than similarity as defined only along the dimension of interest. The morph faces developed and tested in Experiments 2 and 3 provide a set of faces that are controlled in this manner.[6] Aside from being well controlled, these stimuli could be used in situations requiring numerical specification of face attributes, such as when differentiating between prototype and exemplar theories of face categorization (Tanaka, Giles, Kremen, & Simon, 1998; Valentine & Endo, 1992).

In summary, we report evidence that White faces are more sexually dimorphic than Asian faces. Minimizing demand characteristics, this conclusion was based on similarity ratings rather than explicit judgments of gender, which minimizes the contribution of demand characteristics to participant ratings. Furthermore, this conclusion was based on the ratings of a mixed Asian/White population of participants, and we did not find any evidence of own-race bias. This provides an example of correlation between face attributes for a random selection of faces, and such unintended correlations may pose a problem for research designed to address a single attribute. For example, a study designed to reach conclusions about race may report a result that, in truth, reflects unintended variations in gender. The stimuli and techniques reported in Experiments 2 and 3 can be used to avoid the pitfalls of such unintended correlations between race and gender. Furthermore, they can be used to test novel hypotheses that arise from our sexual dimorphism conclusions. For instance, we developed a theory of own-race bias in which gender is a race-specific cue, and we also developed an alternative account of attractiveness differences for faces of different races. More generally, these stimuli they may be of use in a wide variety of studies because they are well characterized along the dimensions of race and gender.

[6] In terms of the applicability of these stimuli, it cannot be overlooked that the morph faces might be perceived as unnatural. For instance, some of these faces correspond to positions in the face-gender dimensions that may rarely, if ever, occur in reality. However, the race/gender ratings from Experiment 3 suggest that participants readily perceived the morph faces in a manner similar to real faces.

References

Baudouin, J. Y., & Tiberghien, G. (2002). Gender is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 362–365. doi:10.1037/0278-7393.28.2.362

Beier, T., & Neely, S. (1992). Feature-based image metamorphosis. *ACM SIGGRAPH Computer Graphics, 26,* 35–42. doi:10.1145/142920.134003

Brown, E., & Perrett, D. (1993). What gives a face its gender? *Perception, 22,* 829–840. doi:10.1068/p220829

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 41,* 510–520.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23,* 803–822. doi:10.1068/p230803

Bülthoff, I., & Newell, F. N. (2005). *Categorical perception of gender: No evidence for unfamiliar faces* (Tech. Rep. No. No. 094). Tübingen, Germany: Max Planck Institute for Biological Cybernetics.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153–176. doi:10.1068/p220153

Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science, 9,* 476–483. doi:10.1111/1467-9280.00088

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin & Review, 11,* 735–741. doi:10.3758/BF03196628

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition, 8,* 237–262. doi:10.1080/13506280042000072

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology, 30,* 527–554. doi:10.1146/annurev.ps.30.020179.002523

Etcoff, N. L. (1986). The neuropsychology of emotional expression. In G. Goldstein & R. E. Tarter (Eds.), *Advances in clinical neuropsychology* (pp. 127–179). New York: Springer. doi:10.1007/978-1-4613-2211-5_5

Finklea, K. M. (2008). *Investigating inappropriate cue utilization in the own-race bias* (Unpublished doctoral dissertation). University of California San Diego, La Jolla, CA.

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research, 1127,* 80–89. doi:10.1016/j.brainres.2006.09.104

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. C. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science, 24,* 498–506. doi:10.1177/0956797612457783

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., . . . Lee, K. (2009). Two faces of the other-race effect: Recognition and categorization of Caucasian and Chinese faces. *Perception, 38,* 1199–1210. doi:10.1068/p6136

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). "Ain't I a woman?": Towards an intersectional approach to person perception and group-based harms. *Sex Roles, 59,* 392–403. doi:10.1007/s11199-008-9505-4

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102,* 116–131. doi:10.1037/a0025335

Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition, 4,* 59–67. doi:10.1080/713756748

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling (quantitative applications in the social sciences)*. London, UK: Sage.

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception, 31,* 567–578. doi:10.1068/p3315

Lovén, J., Rehnman, J., Wiens, S., Lindholm, T., Peira, N., & Herlitz, A. (2012). Who are you looking at? The influence of face gender on visual attention and memory for own- and other-race faces. *Memory, 20,* 321–331. doi:10.1080/09658211.2012.658064

MacLin, O. H., Peterson, D. J., Hashman, C., & Flach, N. (2009). PsychoPro 2.0: Using multidimensional scaling to examine the perceptual categorization of race. *Behavior Research Methods, 41,* 668–674. doi:10.3758/BRM.41.3.668

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. doi:10.1037/1076-8971.7.1.3

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57. doi:10.1037/0096-3445.115.1.39

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An "other-race effect" for categorizing faces by sex. *Perception, 25,* 669–676. doi:10.1068/p250669

Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1997). Facial distinctiveness and the power of caricatures. *Perception, 26,* 207–223. doi:10.1068/p260207

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. doi:10.1111/j.2044-8309.1981.tb00485.x

Steyvers, M. (1999). Morphing techniques for manipulating face images. *Behavior Research Methods, Instruments & Computers, 31,* 359–369. doi:10.3758/BF03207733

Sturrock, K., & Rocha, J. (2000). A multidimensional scaling stress evaluation table. *Field Methods, 12,* 49–60. doi:10.1177/1525822X0001200104

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42,* 7–67. doi:10.1007/BF02293745

Tanaka, J., Giles, M., Kremen, S., & Simon, V. (1998). Mapping attractor fields in face space: The atypicality bias in face recognition. *Cognition, 68,* 199–220. doi:10.1016/S0010-0277(98)00048-1

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 43,* 161–204. doi:10.1080/14640749108400966

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 44,* 671–703. doi:10.1080/14640749208401305

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review, 15,* 1093–1099. doi:10.3758/PBR.15.6.1093

Received October 1, 2013
Revision received May 27, 2014
Accepted May 28, 2014 ■