The flip side of the other-race coin: They all look *different* to me

Abstract

Better recognition of own-race faces than other-races faces has been attributed to a problem of discrimination (i.e., telling faces apart). The conclusion that `they all look the same to me' is based on studies measuring the perception/memory of highly controlled stimuli, typically involving only one or two images of each identity. We hypothesized that such studies underestimate the challenge involved in recognizing other-race faces because in the real world an individual's appearance varies in a number of ways (e.g., lighting, expression, hairstyle), reducing the utility of relying on pictorial cues to identity. In two experiments Caucasian and East Asian participants completed a perceptual sorting task in which they were asked to sort 40 photographs of two identities into piles such that each pile contained all photographs of a single identity. Participants perceived more identities when sorting other-race faces than own-race faces, both when sorting celebrity (Experiment 1) and non-celebrity (Experiment 2) faces, suggesting that in the real world `they all look different to me'. We discuss these results in light of models in which each identity is represented as a region in a multi-dimensional face space; we argue that this region is smaller for other-race than own-race faces.

Keywords: face perception, social cognition, visual perception

The flip side of the other-race coin: They all look *different* to me

People are worse at recognizing and discriminating other-race faces than own-race faces (see Bothwell, Brigham & Malpass, 1989; Meissner & Brigham, 2001 for reviews)[1]. This other-race effect (ORE) is robust across a range of methodologies: recognition tests, in which participants' ability to discriminate between previously seen faces and novel faces is measured (e.g., Golby, Gabrieli, Chiao & Eberhardt, 2001; MacLin & Malpass, 2001; Wright, Boyd & Tredoux, 2003); identity tasks in which participants locate a target face in an identity line-up from memory (e.g., Meissner, Tredoux, Parker & MacLin, 2005; Jackiw, Arbuthnott, Pfeifer, Marcon & Meissner, 2008; Evans, Marcon & Meissner, 2009); discrimination tasks that involve making same/different judgments about pairs of faces (e.g., Walker & Tanaka, 2003; Mondloch et al., 2010); and sequential matching tasks (e.g., Lindsay, Jack & Christian, 1991; Tanaka, Kiefer & Buklach, 2004; Rhodes, Hayward & Winkler, 2006).

In addition to providing insights about the role of experience in the development of perceptual expertise (see Tanaka, Heptonstall & Hegan, 2013; Kelly et al., 2007), this phenomenon has important practical implications. Difficulty in recognizing other-race individuals leads to embarrassment when adults fail to recognize familiar individuals in social or professional contexts, and has led to numerous false incarcerations based on erroneous eyewitness testimony (reviewed in Hugenberg, Young, Bernstein & Sacco, 2010). Understanding the mechanisms underlying the effect is essential.

---

[1] We are using the terms own race and other race to be consistent with the literature but we recognise that these are perceptual/cognitive terms and not biological categories.

The ORE has been framed as a problem with individuating (discriminating between) other-race faces, consistent with Feingold's claim that "to the uninitiated American, all Asiatics looks alike, while to the Asiatic, all White men look alike" (1914, p50; see also Meissner & Brigham, 2001; Vizioli, Rousselet & Caldara, 2010; but see Goldstein, 1979). When asked to recall faces from memory, participants typically make fewer hits (correctly identifying a previously seen face as familiar) and more false alarms (incorrectly identifying a novel face as familiar) for other-race faces, compared to own-race faces (Meissner & Brigham, 2001). A higher false alarm rate suggests that other-race faces "all look the same" because the viewer is less sensitive to differences between individuals from that racial group. Consistent with this hypothesis, other-race faces are judged to look more similar to each other than are own-race faces when presented in pairs (Byatt & Rhodes, 2004; Papesh & Goldinger, 2010). In fact a number of journal articles investigating the ORE even have the phrase "they/we all look the same" in their titles (Johnson & Fredrickson, 2005; Ackerman et al., 2006; Wilson & Hugenberg, 2010).

Poor discrimination and recognition of other-race faces is predicted by Valentine's model (Valentine, 1991), according to which each individual face is represented as a unique point in a multidimensional face space. The location of each face is determined by its values on the dimensions underlying face space, along which faces vary (e.g., distance between the eyes, nose length). The dimensions of face space are refined through perceptual experience to represent the facial properties that are optimal for discriminating identities from highly familiar categories (see O'Toole & Natu, 2013 for a discussion); own-race faces are distributed in the central region of face space whereas other-race faces are tightly clustered together in the periphery (Valentine, 1991, also see Fig 1a). This dense clustering of other-race faces is responsible for increased errors when discriminating between other-race identities.

Extensions of Valentine's model take into account an aspect of face recognition that has largely been ignored in the literature (see Burton, 2013)—the fact that representations of each identity can be activated by multiple images; we need, for example, to recognize our neighbor when she dons a pair of sunglasses or applies makeup prior to going out. Voronoi regions (Lewis & Johnston, 1999) and attractor fields (Tanaka, Giles, Kremen & Simon, 1998; also see Tanaka & Corneille, 2007) around each point in face space reflect the range of inputs that are perceived as belonging to a given identity, allowing recognition despite changes in appearance (e.g., in expression, makeup, hairstyle, illumination, or orientation). The size of an identity's attractor field is determined by the density of nearby representations (i.e., by its location in face space) and determines the range of acceptable inputs. Because own-race faces have larger inter-face distances than other-race faces, which are clustered together in the periphery of face space, these models imply that own-race face have larger attractor fields (or Voronoi regions) than other-race faces (see Fig 1a).

In the vast majority of studies investigating the ORE an individual's face is only represented by a single photograph (e.g., Mondloch et al., 2010; Wilson & Hugenberg, 2010; Hancock & Rhodes, 2008) or by a pair of pictures that vary in expression (e.g., Vizioli et al., 2010; Ackerman et al., 2006; Chiroro & Valentine, 1995), viewpoint (e.g., Ellis & Deregowski, 1981; Sporer, Trinkl & Guberova, 2007; Sporer & Horry, 2011) or the camera with which the pictures were taken (e.g., Megreya, White & Burton, 2011; but see Meissner, Susa & Ross, 2013 who varied expression and camera). The ability to recognize that multiple images of another-race face belong to the same person (i.e., the implication of other-race faces having smaller attractor fields) has been ignored.

This is an important oversight: Within-person variability can have a profound effect on one's perception of identity. Even for own-race faces, photos of the same person can be perceived as belonging to different individuals, unless that person is familiar. Jenkins, White, Montfort & Burton (2011) collected 20 photographs of each of two Dutch celebrities. Participants were asked to sort the faces such that all of the photos of the same person were grouped together. Their results were striking: When the faces were familiar (in the Netherlands) most participants correctly sorted the photographs into two identities. However, UK participants who were unfamiliar with the faces perceived more identities (i.e., sorted faces into more piles; Median = 7.5) than the two identities that were present. These findings highlight the difficultly of recognizing unfamiliar identities across natural variation in images.

In the current study we hypothesized that participants would perceive even more identities when completing the sorting task with unfamiliar other-race faces. At first glance, this prediction is counterintuitive; if other-race faces all "look the same" one might expect participants to make fewer piles when sorting other-race faces. However, smaller attractor fields for other-race compared to own-race faces was expected to make recognition of other-race faces across a wide range of natural variation especially hard because even trivial changes might result in an image crossing the boundaries of the identity's relatively small attractor field, resulting in an activation of neighboring identities (see Fig 1b).

To test this hypothesis, in two experiments we asked participants to sort 40 photographs into piles such that each pile contained all of the photographs of one person. Participants were not told that the correct solution was two piles of 20 pictures. In Experiment 1 Caucasian participants sorted photographs of either two Caucasian or two East Asian celebrities; to control for stimulus effects Chinese participants also sorted the Caucasian photographs. In Experiment 2 participants sorted non-celebrity faces and we used a complete design such that both Caucasian and Chinese participants sorted own- and other-race faces. We hypothesized that participants would make more piles (i.e., perceive more identities) when sorting other-race faces then own-race faces.

We also recorded misidentification errors, defined as sorting the two different identities into the same pile. Based on Jenkins et al. (2011) we anticipated very few misidentification errors when participants sorted own-race faces. However, given the predominant view that other-race faces are perceived as more similar than own-race faces, we predicted more misidentification errors for other-race faces than own-race faces.

Method

Participants

Seventy-five participants were included in the final analysis: 25 were East Asian students at XXXXXXXX University, XXXXXXXX (15 female; Mean age = 20.92; SD = 2.74) and 50 were Caucasian students at XXXXXXXX University, XXXXXXXX (45 female; Mean age = 19.48; SD = 1.23). All East Asian and 25 Caucasian participants completed the task with Caucasian faces and 25 Caucasian participants did so with East Asian faces. We aimed to have 25 participants in each condition who were wholly unfamiliar with the identities contained in the sorting card task so we excluded an additional 17 participants who believed they were familiar with the faces. In fact, none of the excluded participants accurately identified the identities contained in the task; all of the identities were misidentified (e.g., as belonging to an American singer or a Japanese actress).

Stimuli

Twenty images of each of two UK celebrities (Holly Willoughby and Fearne Cotton) and two Chinese celebrities (Bingbing Fan and Zhiling Lin) were taken from the Internet via a Google image search. The celebrities were chosen because they were well known in their country of origin, unfamiliar to the participants we tested in other countries and, within each country of origin, physically similar (e.g., hair color, age etc.). For each person we selected the first 20 images in which their face was bigger than 150 pixels in height, displayed in frontal aspect, and not occluded in any way. This resulted in a total of 80 images (20 per identity). The images were cropped so that only the head was displayed (much like a passport photograph) and were changed to grayscale. They were then printed on cards that were 38 x 50 mm in size. A representation of the variability among photographs is shown in Figure 2.

Procedure

The task procedure was also based on Jenkins et al. (2011). Participants were presented with the following written instructions: "*In front of you is a deck of 40 face photos. Your task is to sort the photos by identity, so that photos of the same face are* *grouped together. There is no time limit on this task and you are free to create as many or as few groups as you wish.*" After each participant had completed the card-sorting task they were asked to indicate whether they were familiar with any of the faces. If participants indicated that a face(s) was familiar they were then asked to provide a name(s) or any information about that person (e.g., where they had seen that person). All participants reported minimal experience with other-races.

Results

Number of Perceived Identities

One-sample *t*-tests show that in all three groups the number of perceived identities was significantly greater than the two identities that were present [Caucasian participants sorting Caucasian photos: $t\ (24) = 3.99$, $p = .001$, $d = 1.13$, 95% CI (1.35, 4.25); Caucasian participants sorting East Asian photos: $t\ (24) = 6.96$, $p < .001$, $d = 1.97$, 95% CI (6.30, 11.62); East Asian participants sorting Caucasian photos: $t\ (24) = 8.10$, $p < .001$, $d = 2.29$, 95% CI (8.64, 14.56)].

Participants perceived more identities (i.e., made more piles) when sorting other-race faces compared to own-race faces. As shown in Figure 3, Caucasian participants sorted own-race photos into a mean of 4.8 identities (SD = 3.51; Median = 4; Mode = 2; Range = 2-16). In contrast, they sorted other-race photos into a mean of 11 identities (SD = 6.43; Median = 9; Mode = 7; Range = 4-31) and East Asian participants sorted the Caucasian photos into a mean of 13.6 identities (SD = 7.16; Median = 13; Mode = 13; Range = 1-31).

Variance differed across groups and so we analyzed the data using Mann-Whitney *U* tests. Caucasian participants perceived significantly more identities when sorting the East Asian photos than the Caucasian photos [*U* (25, 25) = 93, $p < .001$, two-tailed, $r = .61$]. Likewise, East Asian participants perceived significantly more identities in the Caucasian photographs than did the Caucasian participants [*U* (25, 25) = 76.5, $p < .001$, two-tailed, $r = .65$]. The two groups sorting other-race faces made a similar number of piles.

Misidentification Errors

We analyzed misidentification errors in two ways. First, we compared the number of piles containing two identities when participants sorted own- versus other-race faces. Second, we compared the number of participants who made at least one misidentification error when sorting own- versus other-race faces. The number of misidentification errors was higher when participants sorted other-race faces (see Table 1), although this effect was limited to Caucasian participants. Mann-Whitney *U* tests confirmed that the mean number of misidentification errors (number of piles with two identities) was higher when Caucasian participants sorted other-race faces (M = 3.0) than own-race faces (M = 1.8) [*U* (25, 25) = 192.50, $p = .02$, two-tailed, $r = .33$]. This is a significant finding because making more piles for other-race faces (see above) reduces the number of faces in each pile and, consequently, reduces the chance probability of misidentification errors. However, there was no significant difference in the number of misidentification errors for East Asian and Caucasian participants sorting the Caucasian photographs [*U* (25, 25) = 265.5, $p = 0.35$, two-tailed, $r = .13$].

[Table 1 here]

For Caucasian participants there was a significant association between the race of the faces they were sorting and the number of people who made at least one misidentification error [χ^2 (1) = 5.71, p = .02, ϕ = .11]; whereas 92% of Caucasian participants made at least one misidentification error for other-race faces only 64% made at least one for own-race faces. However, for Caucasian photographs, there was no significant association between the race of participant (East Asian versus Caucasian) and the number who made at least one misidentification error [χ^2 (1) = 2.60, p = .11, ϕ = .05]. Eighty-four percent of Chinese participants made at least one misidentification error. Taken together our results provide partial support for misidentification errors being more likely when sorting other-race faces.

Experiment 2

The findings from Experiment 1 suggest that within-person variability in appearance affects our perception of identity more for other-race faces than own-race faces. However, two characteristics of our design limit the generalizability of our results. Firstly, our stimuli were images of celebrities. Variability in appearance may be greater for celebrities than for people in the general population; if so, then our findings may exaggerate the influence of face race on recognition. Secondly, our choice of stimuli did not allow us to have a complete design; we did not have a condition in which East Asian participants sorted unfamiliar East Asian faces. From a purely practical perspective, we were extremely unlikely to find many East Asian participants for whom the Chinese celebrities were unfamiliar.

We addressed each of these concerns in Experiment 2 in which both East Asian and Canadian participants sorted own- and other-race face photos by identity. All photographs were of non-celebrities and, to increase generalizability, two face pairs were used for each race. Thus, Experiment 2 incorporated a complete design and extended our work to new, non-famous identities.

Method

Participants

We tested a total of 80 participants: 40 were East Asian students at XXXXXXXX University, XXXXXXXX (28 female; Mean age = 22; SD = 1.89) and 40 were Caucasian students at XXXXXXXX University, XXXXXXXX (37 female; Mean age = 17; SD = 2.20). Twenty East Asian and 20 Caucasian participants sorted East Asian faces (10 sorted each of two face pairs). Twenty East Asian and 20 Caucasian participants sorted Caucasian faces (10 sorted each of two face pairs).

Stimuli and Procedure

We recruited four Caucasian non-celebrity models and four East Asian non-celebrity models each of whom allowed us access to their pictures via social media (e.g., Facebook and QQ space). All models were young adult females. We paired up the models from each race, resulting in two Caucasian pairs and two East Asian pairs. The models within each pair were of a similar age and had a similar hair color. We selected the first 20 images from each model's social media Webpage where their face was bigger than 150 pixels in height, displayed a roughly frontal aspect, and not occluded in any way. We also tried to ensure that all of the photographs were taken on different days. As in Experiment 1, the images were cropped, changed to greyscale, printed on cards that were 38 x 50 mm in size, and grouped such that each participant was given a pile of 40 photographs—20 per each of two identities.

Results and Discussion

Number of Perceived Identities

Preliminary analyses revealed no effect of stimulus pair, all *p*s > .20, regardless of whether photographs were sorted by Caucasian or East Asian participants. All subsequent analyses are collapsed across face pairs. As in Experiment 1, in every condition the number of perceived identities was significantly greater than the two identities that were present (Caucasian participants sorting own and other-race faces, t (19) = 4.32, p < .001, d = 1.34, 95% CI (4.55; 9.35); t (19) = 6.68, p < .001, d = 2.07, 95% CI (7.01; 11.59); East Asian participants sorting own and other-race faces, t (19) = 4.70, p < .001, d = 1.46, 95% CI (4.91; 9.59); t (19) = 6.70, p < .001, d = 2.09), 95% CI (7.78; 13.02).

As shown in Figure 4, Caucasian participants sorted own-race photos into an mean of 7.0 identities, (SD = 5.12; Median = 5; Mode = 2; Range = 2-20) and other-race photos into an mean of 9.3 identities, (SD = 4.89; Median = 8.5; Mode = 6; Range = 2-18). East Asian participants sorted the own-race photos into an mean of 7.3 identities, (SD = 4.99; Median = 5; Mode = 5; Range = 2-20) and other-race photos into an mean of 10.4 identities, (SD = 5.6; Median = 10; Mode = 11; Range = 4-22).

We conducted a 2 (participant race: Caucasian vs. East Asian) × 2 (face race: own vs. other race) between-subjects ANOVA. We found a significant main effect of face race ($F_{1, 76}$= 5.68, p = .020, η_p^2 = .070). As in Experiment 1, participants perceived significantly more identities for other-race faces (M = 9.85, SE = 0.816) than own-race faces (M = 7.1, SE = 0.816). The main effect of participant race and the interaction between face race and participant race were both nonsignificant, *p*s > .50.

Misidentification Errors

Unlike Experiment 1, we did not observe any effect of face race on misidentification errors. A 2 (participant race: Caucasian vs. East Asian) × 2 (face race: own vs. other race) between-subjects ANOVA revealed no significant effects, all *p*s > .60. As noted in Table 1, misidentification errors were rare in all conditions.

In summary, participants were more accurate in recognizing that multiple photographs belong to the same identity when sorting own-race faces than when sorting other-race faces. This effect was observed regardless of whether participants were sorting celebrity (Experiment 1) or non-famous identities (Experiment 2). Furthermore, there was only minimal evidence of an impaired ability to discriminate among different other-race identities; although Caucasian participants made more misidentification errors when sorting East Asian photographs compared to Caucasian photographs in Experiment 1, this effect was not replicated in Experiment 2 and in neither experiment did East Asians make more misidentification errors than Caucasians when sorting Caucasian photographs. Collectively, these findings draw attention to the flip side of the other-race effect:

recognizing facial identity despite natural variation in appearance. The implication of these findings is considered in the General Discussion.

General Discussion

Our findings replicate past research, which found that when faces are unfamiliar, pictures of the same person appear to belong to several distinct identities (Jenkins et al., 2011). Most notably, we provide the first evidence that within-person variability affects identity perception of other-race faces even more than it affects identity perception of own-race faces. Our participants perceived more identities when they were sorting other-race faces compared to own-race faces, and this was true for both celebrities and non-celebrities. Whereas research based on perceptual expertise emphasizes the effect of experience on discrimination and recognition (Meissner & Brigham, 2001; Tanaka et al., 2013), our results suggest that experience with a face category also influences perceivers' ability to extract identity information across multiple images, even in a perceptual task in which there are no memory demands (see Sporer et al., 2007; Sporer & Horry, 2011; Meissner et al., 2013 for similar findings in tasks that included memory demands and only two images of each identity).

Our findings cannot be explained by physiognomic differences between East Asian and Caucasian faces (i.e., to pictures from one race being easier to sort); East Asian and Caucasian participants sorted the same faces differently. Perceivers' difficulty in recognizing unfamiliar faces across natural variation in a person's appearance (e.g., changes in lighting and viewpoint) may be the result of unfamiliar face representations being based more heavily on lower-level image properties (e.g., Bruce, 1982; Bruce, Henderson, Newman & Burton, 2001; Burton, Wilson, Cowan & Bruce, 1999; see Young & Bruce, 2011 for a discussion; see Hancock, Bruce & Burton, 2000 for a review), properties that vary across images even when identity is held constant. When recognizing identities across changes in pose, expression, lighting, age or hairstyle, perceivers cannot rely on pictorial cues; rather, they must extract structural information that allows identity matching despite changes in appearance (Bruce & Young, 1986). Our findings are consistent with evidence that adults' ability to extract such structural information from unfamiliar other-race faces is impaired relative to own-race faces (Sporer & Horry, 2011; see also Ellis & Deregowski, 1981).

Valentine's (1991) influential model in which faces are conceptualized as single points in multidimensional face space cannot account for our findings. That model does account for difficulty in discriminating faces from categories with which observers have minimal experience (e.g. other-race and other-age faces); because observers lack sensitivity to differences among these faces (characterized as tightly clustered points in this multidimensional space) one identity is easily mistaken for another (e.g., Byatt & Rhodes, 2004; see Young, Hugenberg, Bernstein & Sacco, 2012 for a review). However, building on Valentine's model, Lewis and Johnston (1999) and Tanaka et al. (1998) characterize each face's representation as a region rather than a single point. The distance between contiguous points determines the size of the region (Voronoi cell, attractor field, respectively) associated with each identity; thus faces in a densely populated location in face space will have smaller attractor fields. Neighboring regions in face space will compete with each other when an ambiguous incoming image is similar to both regions. This could result in two things: i) pictures of different people being incorporated into the same region, or ii) pictures of the same person being separated into different regions. Thus, it is because other-race faces are more similar perceptually to one another than own-race faces that they have smaller attractor fields and, consequently, are more difficult to recognize across natural variation in appearance.

Theories of familiar face recognition (e.g., Jenkins & Burton, 2011; Jenkins et al., 2011; Burton, Jenkins, Hancock & White, 2005) imply that familiarity with a person defines the variability that will be incorporated into their representation. Previous behavioral findings suggest that expert performance with familiar faces does not generalize to unfamiliar faces (e.g., Burton et al., 1999, Bruce et al., 2001). Participants perceive multiple identities in a set of photos of a single person despite knowing hundreds of people of the same race and age (Jenkins et al., 2011). Furthermore, whereas telling participants that only two identities were present in a pile of 40 photographs improved performance with new images of those identities in a subsequent same/different matching task, no improvement was observed for new identities (Andrews, Jenkins, Cursiter, & Burton, 2015). Such findings suggest that variability should be understood for each face separately, rather than for faces as a class of object.

Whereas performance differences for familiar versus unfamiliar own-race faces emphasize the importance of experience with particular face, our findings support a role for experience with a face category (another level of familiarity). While acknowledging that there may be qualitative differences between familiar and unfamiliar face processing (see Burton, 2013 for a discussion), our data suggest that experience with multiple faces from a given category influences one’s ability to recognize identities across images of unfamiliar people. Having fewer familiar other-race exemplars stored in memory might result in less knowledge of how individual other-race faces can vary in appearance, limiting our ability to recognize unfamiliar faces in ambient images. It may be only as we become familiar with multiple other-race individuals that we learn more about how any single identity can vary, just as infants and children need to hear a word spoken by many individuals in order to learn it well (Rost & McMurray, 2009; see Watson, Robbins & Best, 2014 for a review and discussion). As more other-race exemplars are incorporated into face space our sensitivity to facial dimensions increases, increasing inter-face spacing, and, consequently, the size of attractor fields.

A similar process may underlie the development of expert face processing during childhood. Given evidence the children's ability to simultaneously rely on multiple dimensions improves after 8 years of age (Nishimura, Maurer, & Gao, 2009) and that children are less sensitive than adults differences along the dimensions of face space (Jeffery et al., 2010; Short, Lee, Fu, & Mondloch, 2014), we suggest that the development of expertise during childhood can be conceptualized as learning multiple examplars resulting in increased sensitivity to the dimensions of face space, leading to larger inter-face distances and, consequently, larger attractor fields associated with unfamiliar identities. Therefore young children, like adults tested with other-race faces, are expected to make more errors when sorting unfamiliar own-race faces than adults.

The practical implications of our results are significant. Jenkins et al. (2011) highlight the significance of within-person variability on the utility of photo identification. Whereas almost any photograph is easily matched to the correct familiar identity, matching photographs of an unfamiliar person is more challenging (see Johnston & Edmonds, 2009 for a review). Based on evidence that other-race faces are judged more similar to each other than own-race faces, the challenge facing airport security officials is thought to be that of discriminating identities. Our results suggest another challenge: recognizing that the person carrying identification is the person on the passport despite changes associated with hairstyle, weight gain/loss, make-up, etc. (also see Meissner et al., 2013). This challenge is exemplified in the case of Suaad Hagi Mohamud, a Canadian who was detained and imprisoned in Nairobi when an airport official concluded that her 4-year-old passport photo was a picture of someone else (Aulakh, 2009, August 10). Previous studies in which participants were trained to recognize a single image of multiple other-race identities (Lebrecht, Pierce, Tarr & Tanaka, 2009) showed that such training is only minimally effective. To the extent that distance between identities in multidimensional face space is correlated with the size of their Voronoi cells or attractor fields, training people with multiple images of each identity may prove to be more useful (Andrews et al., 2015).

In conclusion, our work gives new insights as to why we find recognizing other-race faces so challenging. Whereas prior work emphasizes an impaired ability to discriminate other-race faces (e.g., recognizing that faces belong to different people), we found an impaired ability to recognize an identity across images that incorporate natural variability (e.g., recognizing that faces belong to the same person). We believe that this should be incorporated into new and existing theories of the ORE.

References

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V. Maner, J. K. & Schaller, M. (2006). They all look the same to me (unless they're angry) from out-group homogeneity to out-group heterogeneity. *Psychol. Sci.* 17, 836-840. doi: 10.1111/j.1467-9280.2006.01790.x

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology*, 1-19. doi: 10.1080/17470218.2014.1003949

Aulakh, R. (2009, August 10) Experts vindicate woman trapped in Kenya. *The Star*. Retrieved from http://www.thestar.com/news/gta/2009/08/10/experts_vindicate_woman_trapped_in_kenya.html

Bothwell, R. K., Brigham, J. C. & Malpass, R. S. (1989). Cross-racial identification. *Pers. Soc. Psychol. B.* 15, 19-25. doi: 10.1177/0146167289151002

Bruce, V. (1982). Changing faces: Visual and non- visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105-116. doi: 10.1111/j.2044-8295.1982.tb01795.x

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7*(3), 207. doi: 10.1037//1076-898X.7.3.207

Bruce, V. & Young, A. (1986). Understanding face recognition. *Brit. J. Psychol.* 77, 305-327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive psychology*, *51*(3), 256-284. doi: 10.1016/j.cogpsych.2005.06.003

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248. doi: 10.1111/1467-9280.00144

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Q. J. Exp. Psychol.* 68, 1467-1485. doi: 10.1080/17470218.2013.800125

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

Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Q. J. Exp. Psychol.-A* 48, 879 - 894. doi: 10.1080/14640749508401421

Ellis, H. D. & Deregowski, J. B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *Am. J. Psychol.* 94, 27-35. doi: 10.2307/1422341

Evans, J. R., Marcon, J. L. & Meissner, C. A. (2009). Cross-racial lineup identification: assessing the potential benefits of context reinstatement. *Psychol. Crime Law* 15, 19-28. doi: 10.1080/10683160802047030

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *J. Crim. Law & Police Sci.* 5, 39-51.

Golby, A. J. Gabrieli, J. D. Chiao, J. Y. & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845-850. doi:10.1038/90565

Goldstein, A. G. (1979). Race-related variation of facial features: Anthropometric data I. *B. Psychonomic Soc.* 13, 187-190. doi:10.3758/BF03335055

Hancock, K. J. & Rhodes, G. (2008). Contact, configural coding and the other- race effect in face recognition. *Brit. J. Psychol.* 99, 45-56. doi: 10.1348/000712608X396503

Hancock, P. J., Bruce, V. & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends Cogn. Sci.* 4, 330-337. doi: 10.1016/S1364-6613(00)01519-9

Hugenberg, K., Young, S. G., Bernstein, M. J. & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychol. Rev.* 117, 1168-1187. doi: 10.1037/a0020463

Jackiw, L. B., Arbuthnott, K. D., Pfeifer, J. E., Marcon, J. L. & Meissner, C. A. (2008). Examining the cross-race effect in lineup identification using Caucasian and First Nations samples. *Can. J. Beh. Sci.* 40, 52-5. doi: 10.1037/0008-400x.40.1.52

Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision, 10*(5), 18. doi: 10.1167/10.5.18

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1571), 1671-1683. doi: 10.1098/rstb.2010.0379

Jenkins, R., White, D., Van Montfort, X. & Burton, A. M. (2011). Variability in photos of the same face. *Cognition* 121, 313-323. doi: 10.1016/j.cognition.2011.08.001

Johnson, K. J. & Fredrickson, B. L. (2005). "We all look the same to me" positive emotions eliminate the own-race bias in face recognition. *Psychol. Sci.* 16, 875-881. doi: 10.1111/j.1467-9280.2005.01631.x

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577-596. doi: 10.1080/09658210902976969

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L. & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3- month-old infants. *Infancy* 11, 87-95. doi: 10.1080/15250000709336871

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PloS one, 4*(1), e4215. doi: 10.1371/journal.pone.0004215

Lewis, M. B., & Johnson, R. A. (1999). A unified account of the effects of caricaturing faces. *Visual Cognition, 6*(1), 1-42. doi: 10.1080/713756800

Lindsay, D. S., Jack, P. C. & Christian, M. A. (1991). Other-race face perception. *J. Appl. Psychol.* 76, 587-589. doi: 10.1037/0021-9010.76.4.587

MacLin, O. H. & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychol., Public Pol. L.* 7, 98-118. doi: 10.1037/1076-8971.7.1.98

Megreya, A. M., White, D. & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *Q. J. Exp. Psychol.* 64, 1473-1483. doi: 10.1080/17470218.2011.575228

Meissner, C. A. & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol. Public Pol. L.* 7, 3-35. doi: 10.1037/1076-8971.7.1.3

Meissner, C. A., Susa, K. J. & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Vis. Cogn.* 21, 1287-1305. doi: 10.1080/13506285.2013.832451

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & cognition, 33*(5), 783-792. doi: 10.3758/BF03193074

Mondloch, C.J., Elms, N., Maurer, D., Rhodes, G., Hayward, W., Tanaka, J. & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception* 39, 1065-1085. doi: 10.1068/p6608

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of experimental child psychology, 103*(3), 355-375. doi: 10.1016/j.jecp.2009.02.005

O'Toole, A. J. & Natu, V. (2013). Computational perspectives on the other race effect. *Vis. Cogn.* doi: 10.1080/13506285.2013.803505

Papesh, M. H. & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition* 116, 283-288. doi: 10.1016/j.cognition.2010.05.001

Rhodes, G., Hayward, W.G. & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychon. B. Rev.* 13, 499-505. doi: 10.3758/BF03193876

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science, 12*(2), 339-349. doi: 10.1111/j.1467-7687.2008.00786.x

Short, L. A., Lee, K., Fu, G., & Mondloch, C. J. (2014). Category-specific face prototypes are emerging, but not yet mature, in 5-year-old children. *Journal of experimental child psychology, 126*, 161-177. doi: 10.1016/j.jecp.2014.04.004

Sporer, S. L. & Horry, R. (2011). Pictorial versus structural representations of ingroup and outgroup faces. *J. Cogn. Psychol.* 23, 974-984. doi: 10.1080/20445911.2011.594434

Sporer, S. L., Trinkl, B. & Guberova, E. (2007). Matching faces differences in processing speed of out-group faces by different ethnic groups. *J. Cross Cult. Psychol.* 38, 398-412. doi: 10.1177/0022022107302310.

Tanaka, J. W., Heptonstall, B. & Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Vis. Cogn.* 21, 1183-1201. doi: 10.1080/13506285.2013.826315

Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: further evidence for the attractor field model. *Perception & psychophysics*, *69*(4), 619-627. doi: 10.3758/BF03193919

Tanaka, J. W., Giles, M., Kremen, S., & Simon, V. (1998). Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition*, *68*(3), 199-220. doi:10.1016/S0010-0277(98)00048-1

Tanaka, J. W., Kiefer, M. & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition* 93, B1-B9. doi: 10.1016/j.cognition.2003.09.011

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol.* 43A, 161-204. doi: 10.1080/14640749108400966

Vizioli, L., Rousselet, G. A. & Caldara, R. (2010). Neural repetition suppression to identity is abolished by other-race faces. *P. Natl. A. Sci. USA*, 107, 20081-20086. doi: 10.1073/pnas.1005751107

Walker, P. M. & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception* 32, 1117-1126. doi: 10.1068/p5098

Watson, T. L., Robbins, R. A., & Best, C. T. (2014). Infant perceptual development for faces and spoken words: An integrated approach. *Developmental psychobiology*, *56*(7), 1454-1481. doi: 10.1002/dev.21243

Wilson, J. P. & Hugenberg, K. (2010). When under threat, we all look the same: Distinctiveness threat induces ingroup homogeneity in face memory. *J. Exp. Soc. Psychol.* 46, 1004-1010. doi: 10.1016/j.jesp.2010.07.005

Wright, D. B., Boyd, C. E. & Tredoux, C. G. (2003). Inter- racial contact and the own-race bias for face recognition in South Africa and England. *Appl. Cognitive Psychol.* 17, 365-373. doi: 10.1002/acp.898

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, *102*(4), 959-974. doi: 10.1111/j.2044-8295.2011.02045.x

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and Motivation in Face Recognition A Critical Review of Theories of the Cross-Race Effect. *Personality and Social Psychology Review, 16*(2), 116-142. doi: 10.1177/1088868311418987

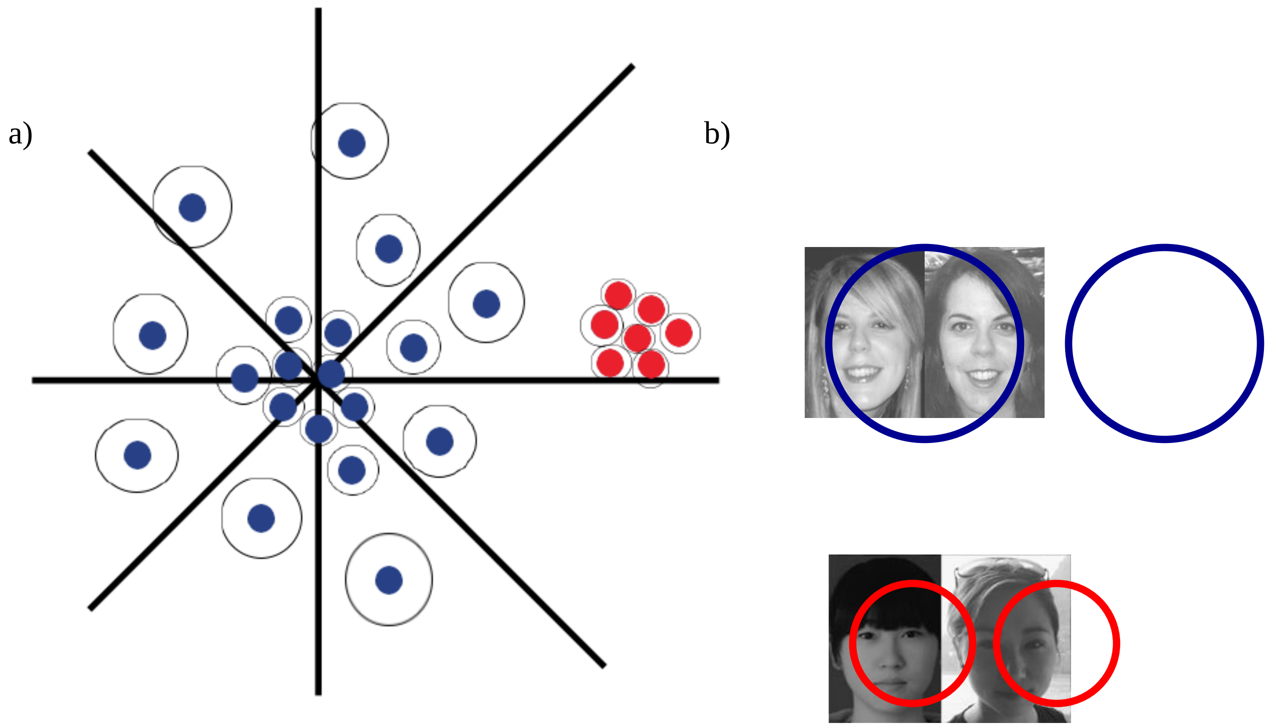

Figure 1

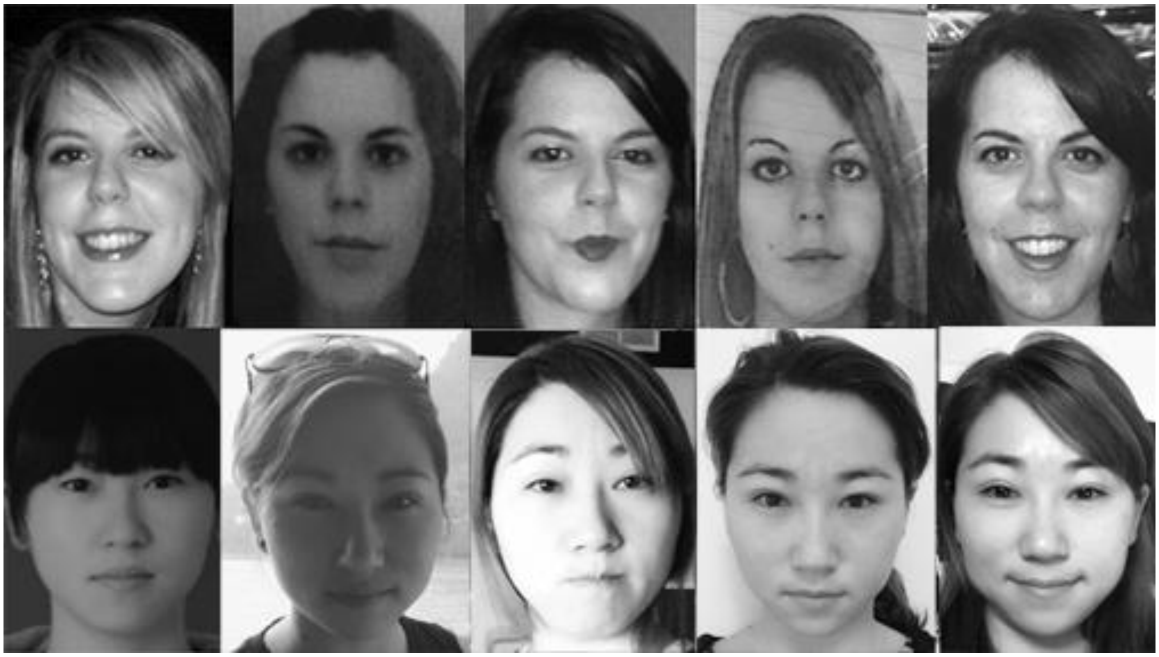

Figure 2

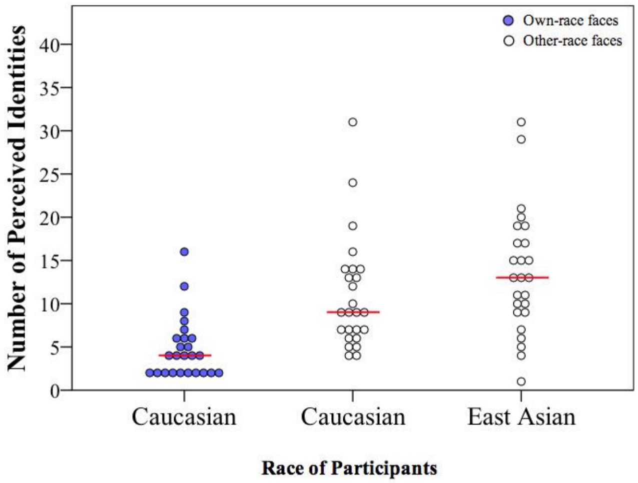
Own-race faces
Other-race faces
Number of Perceived Identities
40
35
30
25
20
15
10
5
0
Caucasian
Caucasian
East Asian
Race of Participants

Figure 3

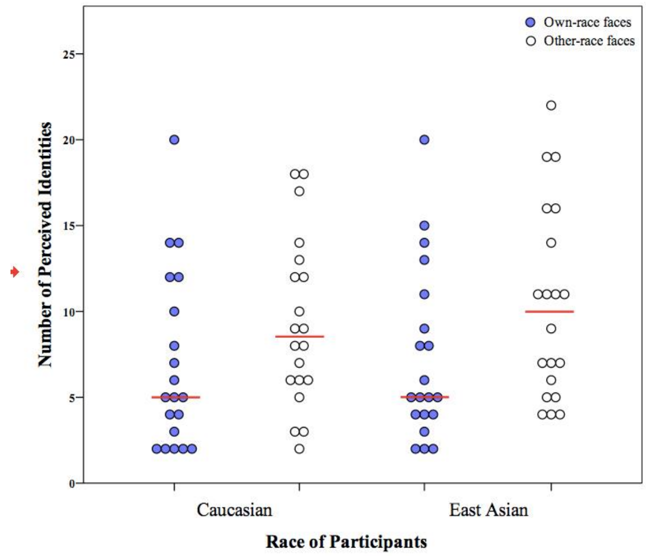
Own-race faces
Other-race faces
Number of Perceived Identities
25
20
15
10
5
0
Caucasian
East Asian
Race of Participants

Figure 4

Figure Captions

Figure 1. a) A representation of Valentine's (1991) face space, where each dot represents an individual face and the circles around each dot represent the attractor fields. Own-race faces (blue dots) fall in the center of face-space and have relatively large attractor fields whereas other-race faces (red dots) are tightly clustered together in the periphery of face-space with relatively small attractor fields. b) Each circle represents a face and its associated attractor field in face-space. Top row: Own-race faces are further apart and the attractor field is bigger. Two pictures of the same Caucasian identity both fall within the same attractor field; therefore they are perceived as belonging to the same person. Bottom row: Other-race faces are closer together and the size of the attractor field is smaller. Two pictures of the same East Asian identity overlap with two attractor fields; therefore they are perceived as belonging to two distinct identities.

Figure 2. Five pictures of XXXXXXXX (top) and XXXXXXXX (bottom). We are unable to show the photographs from our experiment for copyright reasons.

Figure 3. The number of perceived identities for Caucasian and East Asian participants sorting Caucasian or East Asian faces. Each dot represents the number of perceived identities for an individual participant. The red line represents the median number of perceived identities.

Figure 4. The number of perceived identities for Caucasian and East Asian participants sorting own- and other-race faces. Each dot represents the number of perceived identities for an individual participant and the red line depicts the median.

Table 1

Descriptive Statistics for the Number of Misidentification Errors in Experiment 1 and 2.

	M	*SD*	Range	Median	Mode
Experiment 1					
CA sorting Own	1.8	1.88	0-6	1	0
CA sorting Other	3.0	1.70	0-6	3	3
EA sorting Other	2.2	2.02	0-6	1	1
Experiment 2					
CA sorting Own	1.4	1.43	0-5	1	0
CA sorting Other	1.6	1.67	0-5	1	0
EA sorting Own	1.2	1.61	0-5	0	1
EA sorting Other	1.4	1.47	0-4	1	0

Note. CA = Caucasian; EA = East Asian