Learning context and the other-race effect: Strategies for improving face recognition

Jacqueline G. Cavazos[*], **Eilidh Noyes**, **Alice J. O'Toole**

The University of Texas at Dallas, USA

Abstract

People recognize faces of their own race more accurately than faces of other races—a phenomenon known as the "Other-Race Effect" (ORE). Previous studies show that training with multiple variable images improves face recognition. Building on multi-image training, we take a novel approach to improving own- and other-race face recognition by testing the role of learning context on accuracy. Learning context was either *contiguous*, with multiple images of each identity seen in sequence, or *distributed*, with multiple images of an identity randomly interspersed among different identities. In two experiments, East Asian and Caucasian participants learned own- and other-races faces either in a contiguous or distributed order. In Experiment 1, people learned each identity from four highly variable face images. In Experiment 2, identities were learned from one image, repeated four times. In both experiments we found a robust other-race effect. The effect of learning context, however, differed depending on the variability of the learned images. The distributed presentation yielded better recognition when people learned from single repeated images (Exp. 1), but not when they learned from multiple variable images (Exp. 2). Overall, performance was better with multiple-image training than repeated single image training. We conclude that multiple-image training and distributed learning can both improve recognition accuracy, but via distinct processes. The former broadens perceptual tolerance for image variation from a face, when there are diverse images available to learn. The latter effectively strengthens the representation of *differences* among similar faces, when there is only a single learning image.

Keywords

Face recognition; Other-race effect; Training; Learning context

1. Introduction

The other-race effect (ORE) is the tendency for people to recognize faces of their "own" race more accurately than faces of "other" races (cf., Meissner & Brigham, 2001). Often characterized as an inability to discriminate or individualize other-race faces, the ORE suggests that there are clear differences in how own- and other-race faces are processed. The ORE has been studied extensively in face perception. Its effects have been examined across a variety of racial and ethnic groups, using multiple paradigms, (Meissner & Brigham,

[*]Corresponding author at: The University of Texas at Dallas, School of Behavioral and Brain Sciences, GR41, 800 W Campbell Road, Richardson, TX 75080-3021, USA. jacqueline.cavazos@utdallas.edu (J.G. Cavazos).

2001), and more recently in face recognition algorithms (Phillips, Jiang, Narvekar, Ayyad, & O'Toole, 2011). Additionally, the ORE has been found in infants as young as three months (Sangrigoli & de Schonen, 2004), in children (Anzures et al., 2014; Pezdek, Blandon-Gitlin, & Moore, 2003; Tham, Bremner, & Hay, 2017), and in non-typically developing populations, such as individuals with schizophrenia (Pinkham et al., 2008) and autism spectrum disorder (Wilson, Palermo, Burton, & Brock, 2011; Yi et al., 2016).

In recent research, there has been new emphasis on ways to improve face recognition (Heron-Delaney et al., 2011; Hugenberg, Miller, & Claypool, 2007; Ritchie & Burton, 2016; Rodríguez, Bortfeld, & Gutiérrez-Osuna, 2008; Tanaka & Pierce, 2009; White, Kemp, Jenkins, & Burton, 2014; Xiao et al., 2015). Several of these studies have explored ways to reduce the ORE. For example, participant awareness of the ORE (Hugenberg et al., 2007) and viewing caricatured images of faces (Rodríguez et al., 2008) reduce the ORE. Furthermore, developmental studies show that perceptual training in infants can prevent an ORE from emerging (Heron-Delaney et al., 2011).

The effects of perceptual training on learning other-race faces have been studied also in children (Xiao et al., 2015) and adults (Tanaka & Pierce, 2009). For example, Xiao et al. (2015) examined the influence of individuation and categorization training of other-race faces for preschool children. Recognition performance was measured in the context of implicit racial bias. When trained to individuate African American faces, Chinese preschoolers demonstrated a reduced implicit bias for other-race faces (they were less likely to categorize an angry racially ambiguous face as African American). These results suggest that individuation training reduces implicit bias, which is a potential step towards reducing the ORE. Tanaka and Pierce (2009) investigated the effect of individuation and categorization training of Hispanic and African American faces on Caucasian adults. Participants in the individuation condition showed marginally better recognition than those in the categorization condition. These results suggest that individuation training may reduce the ORE.

In addition to the focus on individuation training, repetitive and variable-image learning have been tested as factors that may improve recognition. Repeated exposure to a *single image* has been shown to increase recognition accuracy (cf., for a meta-analysis of these effects, Shapiro & Penrod, 1986). Recently, the importance of within-person variability has been considered as a factor in face recognition (Dowsett, Sandford, & Burton, 2016; Jenkins, White, Van Montfort, & Burton, 2011; Murphy, Ipser, Gaigg, & Cook, 2015). The human ability to "see" multiple, variable images of the same person as a single identity has been referred to in the literature as the ability to "tell people together" (Andrews, Jenkins, Cursiter, & Burton, 2015; Jenkins et al., 2011). This contrasts to the oft-cited human ability to "tell faces apart", which is long thought to be the core of human face expertise for faces. It is generally understood now that the former is a characteristic of familiar face perception, whereas the latter applies to both familiar and unfamiliar face processing.

For example, Murphy et al. (2015) studied the importance of image variability by training participants with repeated or variable face images of each identity. Recognition accuracy for participants who trained with unique exemplars was greater than those who simply trained

with the same image of each individual, which was repeated multiple times. Dowsett et al. (2016) also found that learning multiple, variable images increased recognition accuracy. Using a card-sorting task, participants were asked to identify a target identity from a comparison set of 30 images. Performance increased with the number of distinct target images seen. Additionally, Ritchie and Burton (2016) examined face recognition accuracy using low and high image variability. Participants who learned names of unfamiliar, Australian celebrities in the high variability condition (different images taken from a Google search) performed more accurately than those who learned names in the low variability condition (multiple still images extracted from within the same video). Ritchie and Burton (2016) found similar results using a same/different face matching task, such that participants who learned faces from multiple variable images performed more accurately. In combination, the results indicate that multiple images provide examples of within-person variations that benefit recognition generalization to novel images.

Only three studies have used own-and other-race faces to examine the impact of multiple images on recognition (Hayward, Favelle, Oxner, & Chu, 2016; Laurence, Zhou, & Mondloch, 2016; Matthews & Mondloch, 2017). Laurence et al. (2016) used multiple images of own- and other-race faces for a sorting task. Participants sorted photographs of four East Asian and four Caucasian celebrities and non-celebrities based on identity. The results showed that photographs of other-race faces (for both celebrities and non-celebrities) were grouped into *more* identities than own-race faces. These results suggest that part of the ORE may be due to poor tolerance of image variation in perceiving a unitary identity in an other-race face.

Hayward et al. (2016) used a novel name-learning paradigm with multiple images to examine the ORE. Participants learned the names of own-race and other-race faces in two learning sessions. In each learning session, Caucasian and Chinese participants learned names for eight own- and other-race faces using two distinct images for each identity. In the testing phase, participants determined whether names and faces matched. Importantly, different images appeared for both training sessions and at test. Throughout the three phases, participants viewed a total of five diverse images of each identity (two in the first learning, two in the second learning, and one at test). Hayward and colleagues found that own-race faces were learned more rapidly than other-race faces and that the final identification task resulted in a robust ORE.

Matthews and Mondloch (2017) also examined the effect of multi-image training for Caucasian participants learning other-race faces (Black individuals). Using a novel, five-day training paradigm, participants learned: a) six different identities using multiple varying images, or b) 12 different identities using a single image presented repeatedly. Participants also completed a same/different pre- and post-test to measure recognition accuracy. The results indicated that this training paradigm improved recognition accuracy for participants trained using the multiple varying images, but not for participants who learned from single repeated images. These findings provide further support for the benefits of multiple-image training in improving recognition accuracy, even for other-race faces. Additionally, Matthews and Mondloch (2017) found that recognition accuracy improved for learned identities, but not for novel, untrained identities. These results suggest that training benefits

are identity-specific. That is, that training benefits for a previously learned identity do not transfer to improvements in general face recognition accuracy for all (learned and unlearned) faces.

Here, we consider a qualitatively different approach to improving recognition via manipulating the "learning context." Previous work by Roark (2007) with same-race faces examined the effects of two learning conditions: *contiguous* learning (seeing varying images of the same identity in sequence) and *distributed* learning (seeing varying images and identities randomly interspersed). Roark (2007) used an old/new face recognition task, in which participants learned multiple images that varied in viewing angle. These images were high quality photographs taken under constant illumination. Participants saw four images per identity in either a low (the same images repeated four times), medium (two identical images repeated twice) or high (four different images repeated once) image-variability condition. The distributed condition resulted in substantially more accurate recognition than the contiguous condition, and performance increased with increasing image diversity. Therefore, recognition benefited from learning interperson variability (via presentation order) and intra-person variability (via the use of multiple images). As noted by Roark, the effects of distributed learning mirror benefits seen in the word processing literature for "spaced learning," which are consistent with the "multiple trace theory" (Crowder, 1976). This theory posits that distributed learning results in numerous and distinct memory traces, whereas contiguous learning yields a single memory trace.

Our purpose was to determine whether learning context (distributed versus contiguous) and image variability at learning (multiple variable images versus a single repeated image) affect recognition accuracy for own- and other-race faces. In Experiment 1, East Asian and Caucasian participants learned own- and other-race faces from multiple images presented either contiguously or in a distributed presentation order. In Experiment 2, we assessed the influence of learning context for a single repeated learning image. We examined the effects of these variables on recognition accuracy and on response bias. We considered the latter variable as a potentially interesting moderator of how people report recognition decisions for own- and other-race faces.

2. Experiment 1

In Experiment 1, we examined the influence of contiguous and distributed learning contexts on the recognition of own- and-other race faces learned from multiple, variable images. We expected to find a strong ORE. Additionally, we hypothesized that overall recognition accuracy would be greater in the distributed condition as compared to those in the contiguous condition.

2.1. Methods

2.1.1. Participants—One hundred and fifty-four students from the University of Texas at Dallas (UTD) were recruited for this study from the UTD SONA Psychology Research participant pool. Participants provided written informed consent prior to beginning the study and were compensated with a one-hour course credit. Eligible participants had normal or corrected-to-normal vision and self-identified as East Asian or Caucasian. Participant race

was determined through an online demographic survey. Four participants were excluded due to a programming error and one participant due to poor overall accuracy (less than 2.5 standard deviations below the mean). Finally, as determined by the demographic survey, 16 individuals did not self-identify as East Asian or Caucasian and thus were excluded. Therefore, a total of 133 (59 East Asian, 85 Female) participants were included in the analysis. This study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.1.2.   Design—In the learning phase, participants (East Asian or Caucasian) viewed a series of faces (East Asian and Caucasian) and were assigned randomly to one of two learning contexts (contiguous or distributed). In the contiguous condition, images were presented consecutively. In the distributed condition, images were presented randomly. The testing phase was an old/new recognition test with images not seen in the training. The dependent variable was the signal detection measure, d'. We also considered the effects of the independent variables on criterion, C. In addition to the recognition task, a survey was used to gather information about participants' demographics (i.e. gender, age, education) and experience with other-race faces (adapted from Young, 2016) (see Appendix).

2.1.3.   Stimuli—All images were from the Notre Dame Database (Phillips et al., 2007). All face stimuli were cropped to include the person from the neck up and were placed on a 2100×1500-pixel black background.

2.1.3.1.   Learning phase.: The learning phase stimuli consisted of 144 images; four diverse images of 36 identities (18 East Asian, 18 Caucasian). The learning phase was also divided into four race-constant blocks, with each block containing all four images of nine identities. All images in the learning phase had an uncontrolled background and illumination. Images depicted identities posed with a neutral expression (see Fig. 1a/b).

2.1.3.2.   Testing phase.: The testing phase stimuli consisted of 72 identities. Of these 72 identities, 36 were "old" (from the learning phase), and 36 were "new" (novel to the participant). For this phase, images pictured the person in front of a gray background with controlled illumination. The images pictured the person smiling (see Fig. 1c).

2.1.4.   Apparatus—All experimental events were controlled with a 21.5-inch iMac, and programmed with PsychoPy (Peirce, 2007). The demographic survey was administered using Qualtrics survey software (Qualtrics, Provo, UT).

2.2. Procedure

Participants were assigned randomly to either the contiguous or distributed learning context. Written and verbal instructions informed participants that the experiment was designed to test their face recognition abilities. Participants were instructed to pay careful attention to the images in the learning phase, because they would be asked to recall these identities subsequently. No mention of image diversity or different test image was made to participants. In the learning phase, each image appeared for two seconds, followed by a one second interstimulus-interval. In the contiguous context, all four images of each individual

appeared consecutively. In the distributed context, images of individuals were randomly interspersed within the block. In the testing phase, participants responded "old" or "new" using labeled keys on their keyboard. Images in the testing phase remained on the screen until a response was made. A demographic survey was administered at the end of the test phase. Participants who did not identity as East Asian or Caucasian were automatically redirected to the end of the survey and their data were excluded from the analysis.

2.3. Results

2.3.1. Demographic survey—Although most Caucasians reported the United States as their birthplace, only about half of East Asian participants were born in the United States (see Table 1). The majority of participants (irrespective of race) attended high school in the United States (see Table 2). Most Caucasian participants identified English as their native language. However, most East Asian participants reported their native language as either "English and another East Asian language" or only an "East Asian language" (see Table 3). Although, most Caucasian participants reported that both of their parents were born in the United States, most East Asian participants reported that neither of their parents were born in the United States (see Table 4). The same pattern was observed for grandparents.

2.3.2. Accuracy—The data were submitted to a 2 (participant race: East Asian and Caucasian) × 2 (learning context: distributed and contiguous) × 2 (face stimulus race: East Asian and Caucasian) mixed-design analysis of variance with participant race and learning context as between-subjects variables, face stimulus race as a within-subjects variable, and recognition accuracy (d') as the dependent variable.

Contrary to our learning context hypothesis, there was no main effect of learning context, such that recognition accuracy did not differ between contiguous ($M = 1.21$, $SD = 0.53$) and distributed ($M = 1.21$, $SD = 0.45$) learning contexts, $F(1,129) = .14$, $MSE = 0.49$, $p = .71$, $\eta_p^2 = .00001$ (Fig. 2a).

There was no main effect of participant race—recognition accuracy was similar for East Asian participants ($M = 1.27$, $SD = 0.52$) and Caucasian participants ($M = 1.16$, $SD = 0.47$), $F(1,129) = 1.41$, $MSE = 0.49$, $p = .24$, $\eta_p^2 = .01$ (Fig. 2b). There was no main effect for race of stimulus, (East Asian faces, $M = 1.24$, $SD = 0.61$, Caucasian faces, $M = 1.18$, $SD = 0.62$), $F(1,129) = 1.42$, $MSE = 0.26$, $p = .24$, $\eta_p^2 = .01$ (Fig. 2c).

As expected, participants were better at recognizing own-race faces compared to other-race faces, as indicated by the interaction between participant race and stimulus race, $F(1,129) = 9.74$, $MSE = 0.26$, $p = .002$, $\eta_p^2 = .07$ (see Fig. 2d). Notably, the difference in accuracy was most evident for Caucasian and East Asian participants recognizing Caucasian faces.

2.3.3. Criterion—The criterion data were submitted to an ANOVA with the same design. As a standard measure of response bias, this analysis shows how liberal or conservative participants were when making old/new recognition decisions. A liberal response bias indicates a tendency to respond "old", whereas a conservative response bias indicates a tendency to respond "new".

There was a significant effect of learning context, whereby participants in the contiguous learning context ($M = 0.17$, $SD = 0.37$) were more conservative than those in the distributed learning context ($M = -0.06$, $SD = 0.34$), $F(1,129) = 12.91$, $MSE = 0.26$, $p < .0001$, $\eta_p^2 = .09$ (Fig. 3a).

Response bias did not differ for East Asian ($M = -0.04$, $SD = 0.38$) and Caucasian participants ($M = -.07$, $SD = 0.38$), $F(1,129) = .45$, $MSE = 0.26$, $p = .51$, $\eta_p^2 = .003$ (see Fig. 3b). There was a main effect of stimulus race (see Fig. 3c). Participants were more conservative when recognizing Caucasian faces ($M = 0.18$, $SD = 0.45$) than East Asian faces ($M = -0.07$, $SD = 0.40$), $F(1,129) = 45.23$, $MSE = 0.09$, $p < .0001$, $\eta_p^2 = .26$. There was a marginally significant interaction between participant race and face race $F(1,129) = 2.96$, $MSE = 0.09$, $p = .09$, $\eta_p^2 = .02$ (see Fig. 3d). This interaction suggests that although all participants were conservative recognizing Caucasian faces—this effect was especially strong for Caucasian observers. No other interactions were significant.

In a more applied context, these results might suggest a lower tendency to falsely identify Caucasian faces. As we will see, this result was replicated in Experiment 2. We will consider the effects of learning context on response bias in the Discussion.

3. Interim Discussion

In Experiment 1, we examined the effect of two learning contexts (contiguous and distributed) on recognition accuracy for own- and other-race faces. In this experiment, participants learned identities from multiple, highly variable images. As expected, we found a strong ORE. Contrary to our hypothesis, however, distributed learning did not result in higher recognition accuracy than contiguous learning. These results are not entirely consistent with Roark's (2007) finding for a distributed learning context advantage. However, there are two central differences between the two studies. First, our study used own- and other-race faces, whereas Roark (2007) examined own-race recognition. Second, although Roark (2007) used diverse training stimuli, the stimuli varied only in viewpoint, with no difference in illumination or background, even in the high diversity condition (see Fig. 4). In contrast, participants in Experiment 1 viewed multiple, highly variable images to learn identities (see Fig. 1a/b). These images were taken weeks apart and varied in illumination, camera distance, and appearance (e.g., hair style and clothing)—though they did not vary in viewpoint. The additional variability in the present experiment may have resulted in participants experiencing difficulty in spontaneously seeing the different images as the same person (cf. Fig. 1a/b). This might have been the case especially when the images were spaced out in the learning sequence. In other words, participants in Experiment 1, may have struggled at the outset to "tell faces together" (Jenkins et al., 2011) given the image variation. This is unlikely to have occurred with the highly controlled images used by Roark (2007). Therefore, in Experiment 2, we examined learning context using single, repeated images. We hypothesized that when identity equivalence is seen easily (i.e., in the controlled images used by Roark, 2007, or in the identical, repeated images used in Exp. 2), learning context will affect recognition. Experiment 2 also enabled a between-experiment comparison between learning multiple, variable images (Exp. 1) and the same image repeated multiple

times (Exp. 2). This can answer the question of whether the multiple image effects seen in own-race face recognition studies (Dowsett et al., 2016; Jenkins et al., 2011; Longmore, Liu, & Young, 2008; Murphy et al., 2015; Ritchie & Burton, 2016) generalize also to other-race faces in a completely crossed, ORE design.

4. Experiment 2

In Experiment 2, we examined the influence of contiguous and distributed learning context on recognition of own- and-other race faces learned from single repeated images. Again, we expected to find a strong ORE and greater recognition accuracy for the distributed learning context.

4.1. Methods

4.1.1. Participants—Following the same compensation and restrictions in Experiment 1, 155 students from the University of Texas at Dallas (UTD) were recruited using the UTD SONA Psychology Research participant pool. Three participants were excluded due to a programming error and two participants were excluded from the analysis due to poor overall accuracy (less than 2.5 standard deviations below the mean). Finally, ten individuals did not meet the race requirements and thus were also removed from the analysis. Therefore, a total of 140 (65 East Asian, 110 Female) participants were included in the final analysis.

4.1.2. Design—Experiment 2 was identical to Experiment 1, with the exception of the learning stimuli, which were chosen here by randomly selecting one image for each identity (from the four diverse images used in Experiment 1). This image was repeated four times (see Fig. 5). Learning images were counterbalanced across participants to assure that equal numbers of participants learned each identity from each of the available images.

4.2. Results

4.2.1. Demographic survey—Participant demographics in Experiment 1 and 2 were comparable (Tables 1–4) with the exception that in this study, more than half of the East Asian participants were born in the United States.

4.2.2. Accuracy—The data were submitted to a 2 (participant race: East Asian and Caucasian) $\times$ 2 (learning context: distributed and contiguous) $\times$ 2 (face stimulus race: East Asian and Caucasian) mixed-design analysis of variance. Consistent with our main prediction, recognition accuracy was greater for participants in the distributed context ($M = 1.12$, $SD = 0.62$) compared to those in contiguous context ($M = 0.90$, $SD = 0.45$), $F(1,136) = 5.63$, $MSE = 0.60$, $p = .02$, $\eta_p^2 = .04$ (see Fig. 6a).

There was no main effect of participant race, such that recognition accuracy did not differ for East Asian ($M = 0.99$, $SD = 0.49$) and Caucasian ($M = 1.03$, $SD = 0.61$) participants, $F(1,136) = .24$, $MSE = 0.60$, $p = .63$, $\eta_p^2 = .002$ (Fig. 6b). There was also no main effect of face stimuli (East Asian, $M = 1.05$, $SD = 0.62$ and Caucasian, $M = .96$, $SD = 0.66$), although a marginally significant trend showed greater accuracy for East Asian faces $F(1,136) = 3.25$, $MSE = 0.18$, $p = .07$, $\eta_p^2 = .02$ (see Fig. 6c).

As in Experiment 1, a significant interaction between participant race and stimuli race indicated that recognition accuracy was greater for own-race faces than other-race faces, $F(1,136) = 17.10$, $MSE = 0.18$, $p < .0001$, $\eta_p^2 = .11$ (see Fig. 6d). Notably, the difference in accuracy was most evident for Caucasian and East Asian participants recognizing Caucasian faces. No other interactions approached significance.

In summary, we found that the distributed learning context improved recognition accuracy as it did in Roark (2007). It is likely that the lower image diversity in Roark (2007), as compared to Experiment 1, accounts for the difference in learning context effects. When image diversity was eliminated in Experiment 2, we replicated the advantage of distributed over contiguous learning seen in Roark (2007).

4.2.3. Criterion—There was no main effect of learning context such that response bias was similar for contiguous learning ($M = 0.21$, $SD = 0.38$) and distributed learning conditions ($M = 0.13$, $SD = 0.33$), $F(1,136) = 1.44$, $MSE = 0.26$, $p = .23$, $\eta_p^2 = .01$ (see Fig. 7a).

There was no main effect of participant race, such that response bias was similar for East Asian ($M = 0.20$, $SD = 0.43$) and Caucasian participants ($M = .14$, $SD = 0.28$), $F(1,136) = 1.18$, $MSE = 0.26$, $p = .28$, $\eta_p^2 = .01$ (see Fig. 7b). A main effect of race of stimulus indicated that participants were more conservative in their response to Caucasian faces ($M = 0.31$, $SD = 0.40$) than for East Asian faces ($M = 0.04$, $SD = 0.43$), $F(1,136) = 59.40$, $MSE = 0.08$, $p < .0001$, $\eta_p^2 = .30$ (Fig. 7c). Again, there was a marginally significant interaction between participant race and face race, $F(1,135) = 3.34$, $MSE = 0.26$, $p = .07$, $\eta_p^2 = .02$, that again suggested that Caucasian participants were particularly conservative when recognizing faces of their own race (Fig. 7d). No other interactions were significant.

5. Cross experiment analysis

Although not initially designed as a single experiment, we compared recognition accuracy across experiments between training with multiple varying images (Experiment 1) and single repeated images (Experiment 2). Two goals motivated this analysis. First, we wanted to probe for recognition accuracy differences based on image variability (Dowsett et al., 2016; Jenkins et al., 2011). Second, given the influence of learning condition and image variability as seen independently in Exp. 1 and 2, we wanted to examine if learning context and image variation interact to influence face recognition accuracy. The data were analyzed using a 2 (participant race: East Asian and Caucasian) × 2 (learning context: distributed and contiguous) × 2 (face stimulus race: East Asian and Caucasian) × 2 (image variability: multiple varying and single repeated) mixed-design.

5.1. Accuracy

As expected, and evident by a main effect of image variability, recognition accuracy was greater for participants who learned from multiple variable images $M = 1.22$, $SD = 0.49$) than for participants who learned from single repeated images ($M = 1.01$, $SD = 0.56$), $F(1,265) = 10.71$, MSE = 0.55, $p = .001$, $\eta_p^2 = .04$ (see Fig. 8a). Consistent with trace theory,

there was a weak, but non-significant interaction between image variability and learning condition, $F(1,265) = 2.93$, $MSE = 0.55$, $p = .09$, $\eta_p^2 = .01$ (see Fig. 8b). However, given that this four-way comparison was not incorporated in the design of the study, it is possible that this analysis lacked sufficient power to detect an effect.

Additionally, there was a main effect of stimulus race such that recognition accuracy for East Asian faces ($M = 1.16$, $SD = 0.62$) was greater than Caucasian faces ($M = 1.07$, $SD = 0.65$), $F(1,265) = 4.729$, $MSE = 0.22$, $p = .02$, $\eta_p^2 = .02$ (see Fig. 8c). Furthermore, as expected there was a significant interaction between participant race and face race such that recognition accuracy was greater for own compared to other-race faces, $F(1,265) = 25.54$, $MSE = 0.22$, $p = .09$, $\eta_p^2 = .01$ (Fig. 8d). We will speculate on these results in the Discussion. No other results were significant.

5.2. Criterion

Participants, responded more conservatively in the multiple variable ($M = 0.17$, $SD = 0.37$) versus the single repeated image condition ($M = 0.05$, $SD = 0.36$), $F(1,265) = 7.17$, MSE = 0.26, $p = .008$, $\eta_p^2 = .03$ as indicated by a main effect of image variability. There was also no significant interaction between learning condition and image variability, $F(1,265) = 3.01$, MSE = 0.26, $p = .08$, $\eta_p^2 = .01$ for criterion bias.

A main effect of stimulus race indicated that participants responded more conservatively to Caucasian faces ($M = 0.24$, $SD = 0.43$) compared to East Asian faces ($M = -0.16$, $SD = 0.41$), $F(1,265) = 104$, MSE = 0.09, $p < .0001$, $\eta_p^2 = .28$. A main effect of condition demonstrated that response bias was more conservative for contiguous ($M = 0.19$, $SD = 0.38$) compared to distributed learning ($M = 0.04$, $SD = 0.35$), $F(1,265) = 11.623$, MSE = 0.26, $p = .001$, $\eta_p^2 = .04$.

There was also a significant interaction between participant race and race of face such that Caucasian participants responded particularly conservative to their own-race compared to Asian faces, $F(1,265) = 6.38$, MSE = 0.09, $p = .01$, $\eta_p^2 = .02$.

6. Discussion

The primary, novel finding of this study was that learning context (contiguous and distributed) affected recognition accuracy for both own- and other-race faces. Notably, distributed learning yielded greater recognition accuracy, but only when the same learning image was repeated. There was no effect of presentation type when the learning images were highly diverse. Combined with Roark (2007), we conclude that a pre-requisite factor in the utility of distributed learning is the ability of participants to perceive that the images that are repeated in a distributed sequence picture the same person (i.e., limited variability). In other words, the benefits of distributed learning may apply only when the associated images (identical or moderately diverse) are easily "seen together" as a unique identity. As suggested by Roark (2007), one possible explanation of the distributed advantage is the multiple trace theory (Crowder, 1976). This theory suggests that the formation of multiple

memory traces benefits recognition. By this account, distributed presentation provides multiple traces of the experience of seeing a face, whereas contiguous presentation creates a single episodic memory trace for an identity.

The second novel finding was that multi-image learning also benefits recognition accuracy for other-race faces. This complements the benefits of multi-image learning for own-race faces (Dowsett & Burton, 2015; Jenkins et al., 2011; Longmore et al., 2008; Murphy et al., 2015; Ritchie & Burton, 2016), which we replicate here as well. Thus, our findings show that multi-image learning is a promising tool for improving other-race recognition. Data consistent with the utility of multi-image learning for other-race faces were reported in Matthews and Mondloch (2017), but in a design that tested participants of one race, with face stimuli of another race. Results from our cross-experimental analysis demonstrate that this effect applies generally as a cross-race effect. Notably, we found no interaction effects with face and participant race across experiments. This suggests that the benefits of multi-image learning apply equally to faces of own- and other-races, with no indication of qualitatively different effects.

In one sense, it may seem paradoxical to suggest that people struggled at the outset to see the variable images as the same person, and yet they ultimately performed more accurately when the learning images varied than when they learned the same image multiple times. However, the recognition test involved a generalization to a new, quite different image of the face. Thus, learning multiple variable images in the contiguous context may have offset difficulties with the identity grouping, thus allowing for an overall multiple image advantage. These results suggest that learning context may moderate multi-image learning. A more directed test that pits the ability to group identities together and learning context is needed to make stronger conclusions on this point.

Combined, the results of this study elucidate two distinct ways to improve face recognition. These methods apply both to own- and other-race faces. First, image variability improves recognition accuracy by promoting a representation that supports generalization to new images of the same identity. This is a hallmark of familiar face processing that is characterized by a nearly limitless capacity to "tell images of a face together." Second, learning context may promote learning that reinforces memory for identity, presuming enough perceptual support to reinforce the correct memory trace. A more direct test of this latter interpretation would be to vary either the familiarity of a participant with the identities to be learned, or to vary image diversity. We predict that a major constraint on the utility of learning context would be the degree to which participants can group images of identities together at the outset of the learning process. However, it is worth noting that this interpretation is based on a cross-experiment analysis, and should be interpreted with caution. The present study examined the influence of multi-image learning using two independent experiments. Future studies could examine ways to exploit the benefits of *both* learning context and multi-image learning in a single setting. For example, providing participants with an identity cue (e.g. name) might facilitate learning from multi-image sets to more easily group unique identities together in a distributed learning context.

In both experiments, the ORE was most evident for Caucasian faces. One possibility is that reduced experience with Caucasian individuals might have led East Asian participants to find Caucasian faces more difficult to recognize. Although, previous research has demonstrated that individuating experience can influence the ORE (Bukach, Cottle, Ubiwa, & Miller, 2012; Rhodes et al., 2009), we did not directly control for experience in the present study. Results from response biases could provide additional support for the differences found for Caucasian images.

In addition to the effects on recognition accuracy, we also uncovered intriguing response bias differences in these experiments. In both experiments, we found that both East Asian and Caucasian participants showed a strong conservative response bias for Caucasian over East Asian faces. A marginally significant interaction suggested that Caucasian participants were particularly conservative for their own-race in both Experiment 1 and 2. This interaction was significant in the cross-experiment analysis. Taken together, these results suggest that Caucasian faces may be responded to more cautiously. This cautious response bias may have also attributed to the differences in the ORE. A conservative bias might imply that in a judicial setting, individuals would identify Caucasians more cautiously than non-Caucasians. Previous literature has produced inconclusive findings on criterion response bias. Although the majority of studies have found a conservative bi as for own-race faces (Evans, Marcon, & Meissner, 2009; Meissner & Brigham, 2001), others studies have found this bias specifically for Caucasian faces (Jackiw, Arbuthnott, Pfeifer, Marcon, & Meissner, 2008). No study has directly probed criterion bias differences for own- and other-race faces. Moreover, some studies lack a full cross-race design, making it difficult to interpret such biases. Due to the fact that we had no specific hypotheses about criterion, along with the inconsistent findings in the literature, our conclusions here are more speculative. Nonetheless, the results are worth noting, given the effect sizes and their consistency across both experiments and the cross-experiment analysis.

In summary, both learning context and image variability influence face recognition accuracy, but via distinct mechanisms. Each can be used to improve recognition accuracy, but the utility of learning context depends on the ability to "see faces together" at the outset of learning. The results of these experiments have implications for the development of face recognition training programs, especially in applied settings. Importantly, the training mechanisms we studied here are effective for both own- and other-race faces.

Acknowledgements

Appendix

Demographic Survey

Q1 Participant ID

Q2 Please specify your gender

❍ Male

❍ Female

Q3 What is your age in years?

Q4 What is the highest degree you have obtained?

❍ High school or GED

❍ Associates

❍ Bachelors

❍ Masters

❍ PhD, MD or other professional degree

Q5 In what country were you born?

❍ United States

❍ Other

Q6 In what country were you born?

Q7 Have you lived in the United States your entire life?

❍ Yes

❍ No

Q8 How long have you lived in the United States? (in years)

Q9 In what country did you attend high school?

❍ United States

❍ Other

Q10 Where did you attend high school?

Q11 Number of parents born in the United States?

❍ 0

❍ 1

❍ 2

❍ 3

❍ 4

Q12 Number of grandparents born in the United States?

❍ 0

❍ 1

❍ 2

○ 3

○ 4

○ 5

○ 6

Q13 What is your native language/s? (i.e. the language/s you speak at home)

Q14 Are you an international student, or from a country other than the United States?

○ Yes

○ No

Q15 What is your primary racial/ethnic background?

❑ White/Caucasian

❑ African American

❑ East Asian (Thai, Korean, Vietnamese, Chinese, Japanese heritage) American

❑ Other Asian

❑ Hispanic/Latino

❑ Native American

❑ Other

Q16 For international students: What is your primary racial/ethnic background? If you identify with more than one ethnic group, please write all that apply.

❑ White/Caucasian

❑ East Asian (Thai, Vietnamese, Korean, Japanese, Chinese)

❑ Other Asian

❑ Hispanic/Latino

❑ Other

Q17 For East Asian participants: On a scale from 1 (know a little) to 6 (very fluent)

—— Rate your overall English language Ability

—— Rate your overall fluency in an East Asian language?

Q18 On a scale from 1 (Never) to 6 (Always)

—— How often do you watch English language films and TV?

—— How often do you watch Asian language films and TV?

—— How often do you socialize with Caucasians/White people

Q19 Are any of your close friends Caucasian/White?

○ Yes

○ No

Q20 How many close friends are Caucasian/White?

Q21 Are any of your immediate family members Caucasian/White?

○ Yes

○ No

Q22 How many of your immediate family members are Caucasian/White?

Q23 Are any of your extended family members Caucasian/White?

○ Yes

○ No

Q24 How many of your extended family members are Caucasian/White?

Q25 Do you speak any East Asian language?

○ Yes

○ No

Q26 Which ones? How fluent are you in each, on a scale from 1 (know a little) to 6 (very fluent)?

Q27 Have you visited any East Asian countries?

○ Yes

○ No

Q28 Which ones and for how long?

Q29 Regarding social interactions:

—— How often do you socialize with East Asians? On a scale from 1 (Never) to 6 (Always)

Q30 Are any of your close friends East Asians?

○ Yes

○ No

Q31 How many of your close friends are East Asian?

Q32 Are any of your immediate family members East Asians?

○ Yes

○ No

Q33 How many of your immediate family members are East Asian?

Q34 Are any of your extended family members East Asians?

○ Yes

○ No

Q35 How many of your extended family members are East Asians?

References

Andrews S, Jenkins R, Cursiter H, & Burton AM (2015). Telling faces together: Learning new faces through exposure to multiple instances. Quarterly Journal of Experimental Psychology, 68(10), 2041–2050. 10.1080/17470218.2014.1003949.

Anzures G, Kelly DJ, Pascalis O, Quinn PC, Slater AM, de Viviés X, & Lee K (2014). Own- and other-race face identity recognition in children: the effects of pose and feature composition. Developmental Psychology, 50(2), 469–481. 10.1037/a0033166. [PubMed: 23731287]

Bukach CM, Cottle J, Ubiwa J, & Miller J (2012). Individuation experience predicts other-race effects in holistic processing for both Caucasian and Black participants. Cognition, 123(2), 319–324. 10.1016/j.cognition.2012.02.007. [PubMed: 22397819]

Crowder RG (1976). Principles of learning and memory. New York: John Wiley & Sons.

Dowsett AJ, & Burton AM (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. British Journal of Psychology, 106(3), 10.1111/bjop.12103.

Dowsett AJ, Sandford A, & Burton AM (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. The Quarterly Journal of Experimental Psychology, 69(1), 1–10. 10.1080/17470218.2015.1017513. [PubMed: 25671778]

Evans JR, Marcon JL, & Meissner CA (2009). Cross-racial lineup identification: Assessing the potential benefits of context reinstatement. Psychology, Crime and Law, 15(1), 19–28. 10.1080/10683160802047030.

Hayward WG, Favelle SK, Oxner M, & Chu MH (2016). The other-race effect in face learning: Using naturalistic images to investigate face ethnicity effects in a learning paradigm. The Quarterly Journal of Experimental Psychology, 1–7. 10.1080/17470218.2016.1146781.

Heron-Delaney M, Anzures G, Herbert JS, Quinn PC, Slater AM, Tanaka JW, … Pascalis O (2011). Perceptual training prevents the emergence of the other race effect during infancy, 6(5). 10.1371/journal.pone.0019858.

Hugenberg K, Miller J, & Claypool HM (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. Journal of Experimental Social Psychology, 43. 10.1016/j.jesp.2006.02.010.

Jackiw LB, Arbuthnott KD, Pfeifer JE, Marcon JL, & Meissner CA (2008). Examining the cross-race effect in lineup identification using Caucasian and first nations samples. Canadian Journal of Behavioural Science, 40(1), 52–57. 10.1037/0008-400x.40.1.52.

Jenkins R, White D, Van Montfort X, & Burton AM (2011). Variability in photos of the same face. Cognition, 121(3), 313–323. 10.1016/j.cognition.2011.08.001. [PubMed: 21890124]

Laurence S, Zhou X, & Mondloch CJ (2016). The flip side of the other-race coin: They all look different to me. British Journal of Psychology, 107(2), 374–388. 10.1111/bjop.12147. [PubMed: 26366460]

Longmore CA, Liu CH, & Young AW (2008). Learning faces from photographs, 34(1), 77–100. 10.1037/0096-1523.34.1.77.

Matthews CM, & Mondloch CJ (2017). Improving identity matching of newly encountered faces: Effects of multi-image training. Journal of Applied Research in Memory and Cognition, 1–11. 10.1016/j.jarmac.2017.10.005.

Meissner CA, & Brigham JC (2001). Own-race bias in memory for faces a meta-analytic review, 7(1), 3–35. 10.1037//1076-8971.7.1.3.

Murphy J, Ipser A, Gaigg SB, & Cook R (2015). Exemplar variance supports robust learning of facial identity. Journal of Experimental Psychology: Human Perception and Performance, 41(3), 577–581. 10.1037/xhp0000049. [PubMed: 25867504]

Peirce JW (2007). PsychoPy-psychophysics software in python. Journal of Neuroscience Methods, 162(1-2), 8–13. 10.1016/j.jneumeth.2006.11.017. [PubMed: 17254636]

Pezdek K, Blandon-Gitlin I, & Moore C (2003). Children's face recognition memory: More evidence for the cross-race effect. Journal of Applied Psychology, 88(4), 760–763. 10.1037/0021-9010.88.4.760.

Phillips PJ, Scruggs TW, O'Toole AJ, Flynn PJ, Bowyer KW, Schott CL, & Sharpe M (2007). FRVT 2006 and ICE 2006 large-scale results. National Institute of Standards and Technology Internal Report, 7408(NISTIR 7408), 831–846.

Phillips PJ, Jiang F, Narvekar A, Ayyad J, & O'Toole AJ (2011). An other-race effect for face recognition algorithms. ACM Transactions on Applied Perception, 8(11), 10.1145/1870076.1870082 Article.

Pinkham AE, Sasson NJ, Calkins ME, Richard J, Hughett P, Gur RE, & Gur RC (2008). The other-race effect in face processing among African American and Caucasian individuals with schizophrenia, (May), 639–645.

Rhodes G, Ewing L, Hayward WG, Maurer D, Mondloch CJ, & Tanaka JW (2009). Contact and other-race effects in configural and component processing of faces. British Journal of Psychology, 100, 717–728. 10.1348/0007. [PubMed: 19228441]

Ritchie KL, & Burton AM (2016). Learning faces from variability, 218(October). doi: 10.1080/17470218.2015.1136656.

Roark DA (2007). Bridging the gap in face recognition performance: What makes a face familiar? (Doctoral dissertation) University of Texas at Dallas (304766382).

Rodríguez J, Bortfeld H, & Gutiérrez-Osuna R (2008). Reducing the other-race effect through caricatures. In 2008 8th IEEE international conference on automatic face and gesture recognition, FG 2008, 19–21. doi: 10.1109/AFGR.2008.4813398.

Sangrigoli S, & de Schonen S (2004). Recognition of own-race and other-race faces by three-month-old infants. Journal of Child Psychology and Psychiatry and Allied Disciplines, 45(7), 1219–1227. 10.1111/j.1469-7610.2004.00319.x.

Shapiro PN, & Penrod S (1986). Meta-analysis of facial identification studies. Psychological Bulletin, 100(2), 139–156. 10.1037/0033-2909.100.2.139.

Tanaka JW, & Pierce LJ (2009). The neural plasticity of other-race face recognition. Cognitive, Affective, & Behavioral Neuroscience, 9(1), 122–131. 10.3758/CABN.9.1.122.

Tham DSY, Bremner JG, & Hay D (2017). The other-race effect in children from a multiracial population: A cross-cultural comparison. Journal of Experimental Child Psychology, 155, 128–137. 10.1016/j.jecp.2016.11.006. [PubMed: 27965175]

White D, Kemp RI, Jenkins R, & Burton AM (2014). Feedback training for facial image comparison. Psychonomic Bulletin & Review, 21(1), 100–106. 10.3758/s13423-013-0475-3. [PubMed: 23835616]

Wilson CE, Palermo R, Burton AM, & Brock J (2011). Recognition of own- and other-race faces in autism spectrum disorders. The Quarterly Journal of Experimental Psychology, 64(10), 1939–1954. 10.1080/17470218.2011.603052. [PubMed: 21895562]

Xiao WS, Fu G, Quinn PC, Qin J, Tanaka JW, Pascalis O, & Lee K (2015). Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children. Developmental Science, 18(4), 655–663. 10.1111/desc.12241. [PubMed: 25284211]

Yi L, Quinn PC, Fan Y, Huang D, Feng C, Joseph L, … Lee K (2016). Children with Autism Spectrum Disorder scan own-race faces differently from other-race faces. Journal of Experimental Child Psychology, 141, 177–186. 10.1016/j.jeep.2015.09.011. [PubMed: 26435179]

Young LR (2016). Neural correlates of trustworthiness evaluations in cross-cultural interactions. University of Texas at Dallas.
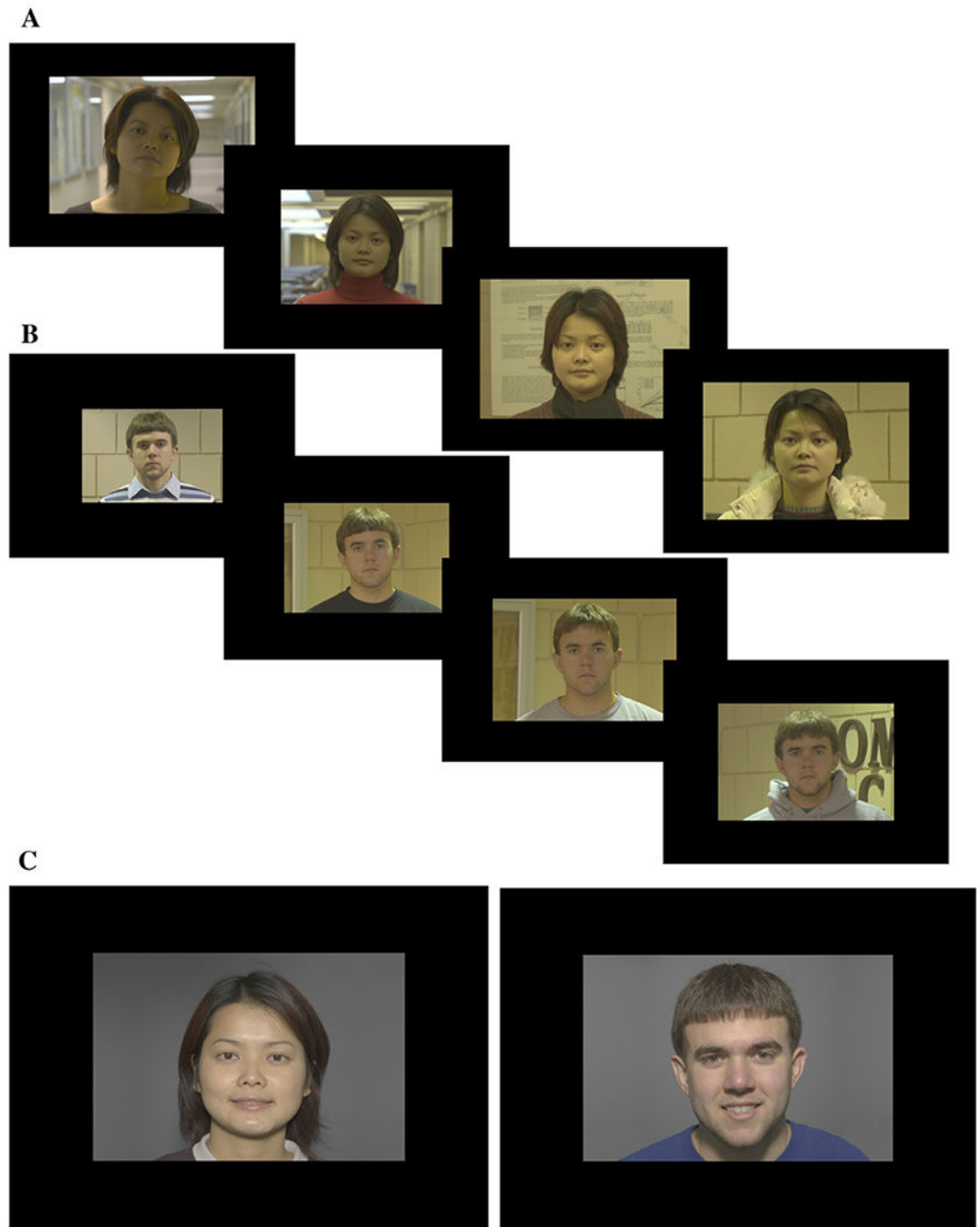
Fig. 1.
Experiment 1 stimuli examples. East Asian (a) and Caucasian (b) learning images. East Asian (left) and Caucasian (right) testing images (c).
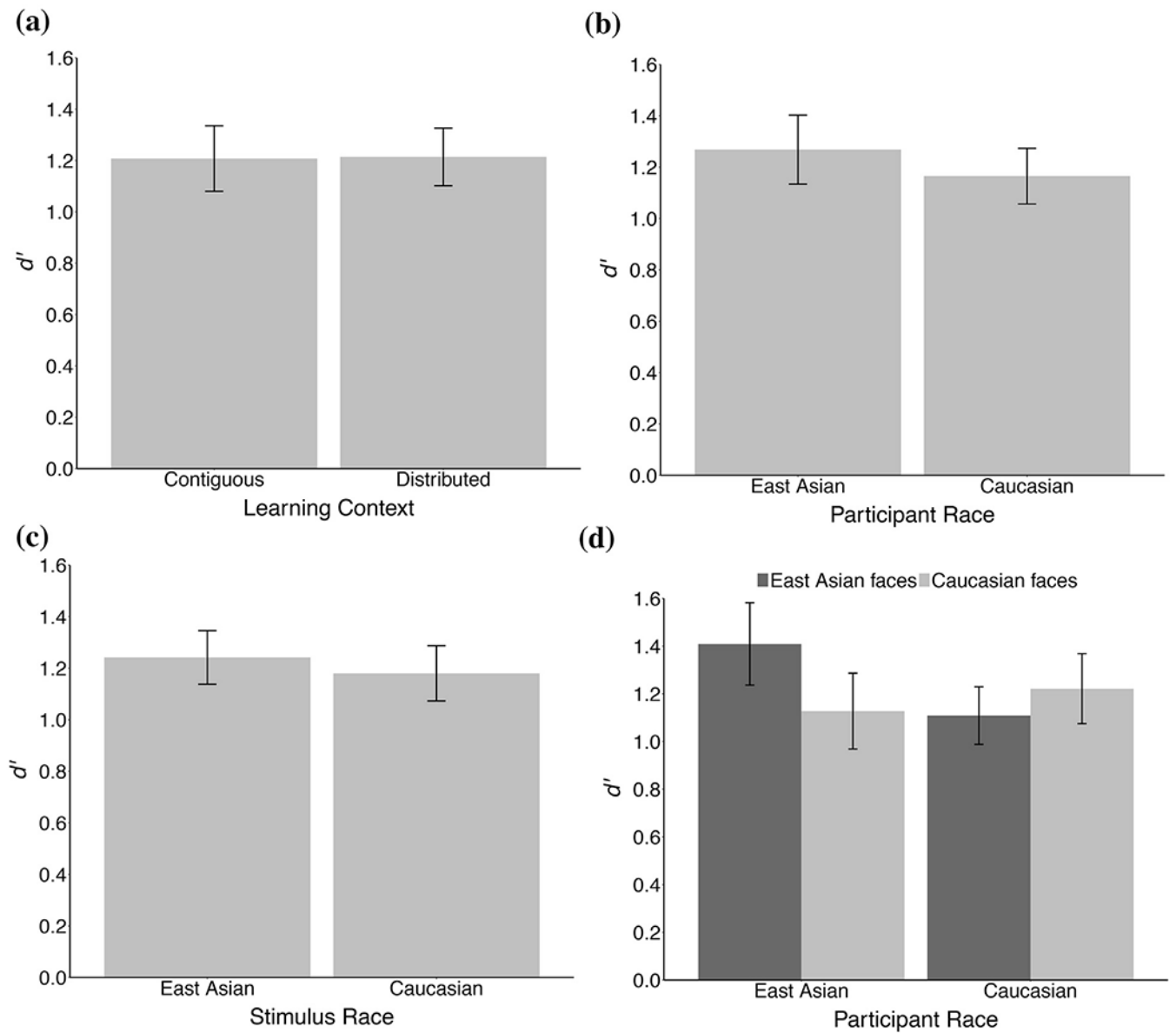
Fig. 2.
No effect of learning context on face recognition accuracy using multiple varying images. Results show that there was no difference in accuracy for learning context (a), participant race (b), or race of face stimuli (c). However, we found a robust other-race effect (d). Error bars show 95% CI.
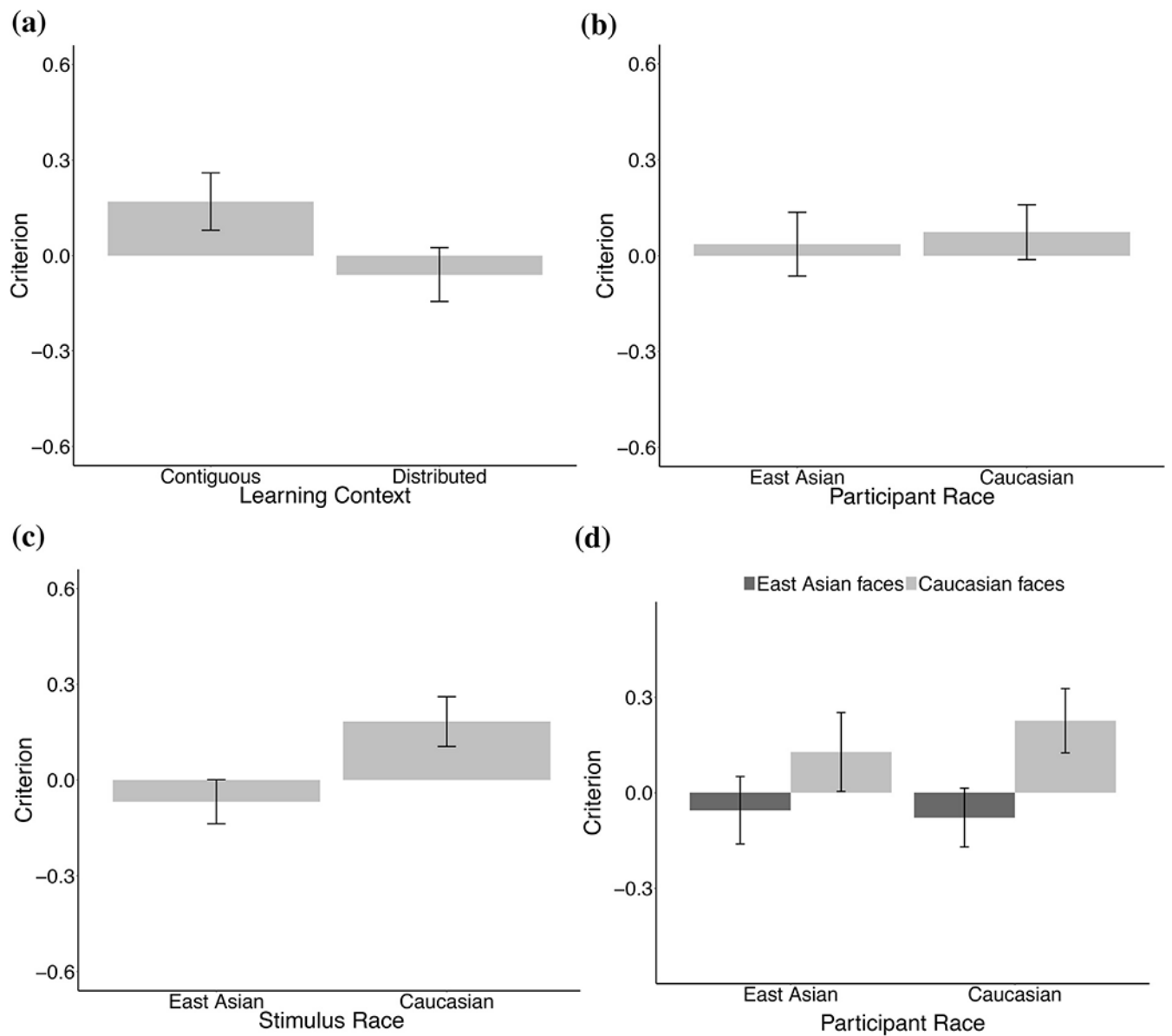
Fig. 3.
Effect of learning context on face recognition criterion bias using multiple varying images. Contiguous learning yielded more conservative responses than distributed learning (a). East Asians and Caucasians did not differ in response bias (b), however participants responded more conservatively to Caucasian faces (c). There was a trend indicating that Caucasians responded more conservatively to Caucasian faces (d). Error bars show 95% CI.

Fig. 4.
Example of diverse learning images in Roark (2007).

Fig. 5.
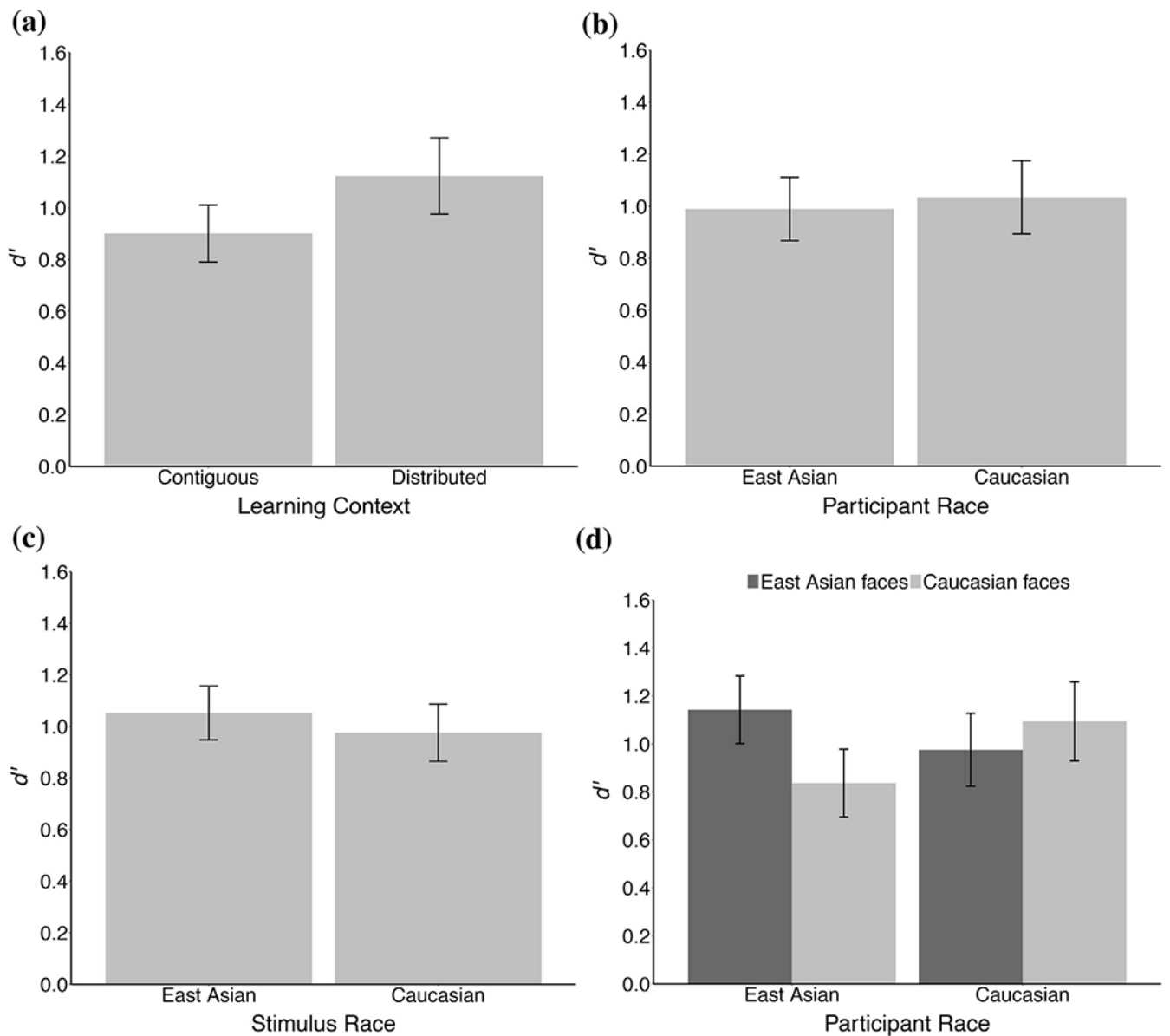Experiment 2 stimuli examples. East Asian (a) and Caucasian (b) learning images.

Fig. 6.
Effect of learning context on face recognition accuracy using single repeated images. Recognition accuracy was greater for contiguous learning than distributed learning (a). East Asians and Caucasians participants did not differ in recognition accuracy (b), however participants responded more conservatively to Caucasian faces (c). There was robust other-race effect (d). Error bars show 95% CI.
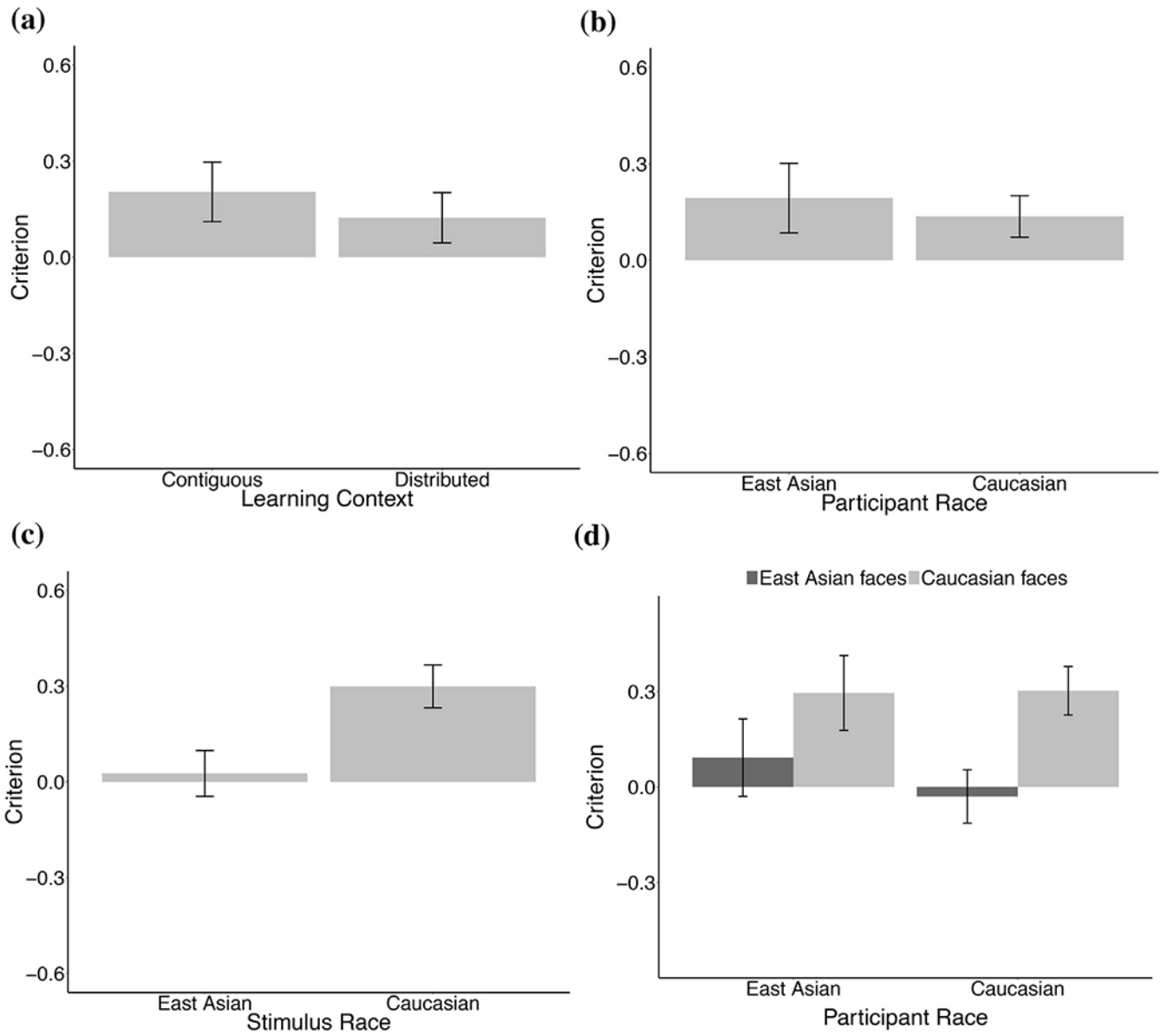
Fig. 7.
No effect of learning context on criterion bias using single repeated images. Results revealed no difference in response bias for learning context (a) or participant race (b). However, there was a greater conservative response bias for Caucasian faces (c) and a trend indicating that Caucasians responded more conservatively to Caucasian faces (d). Error bars show 95% CI.
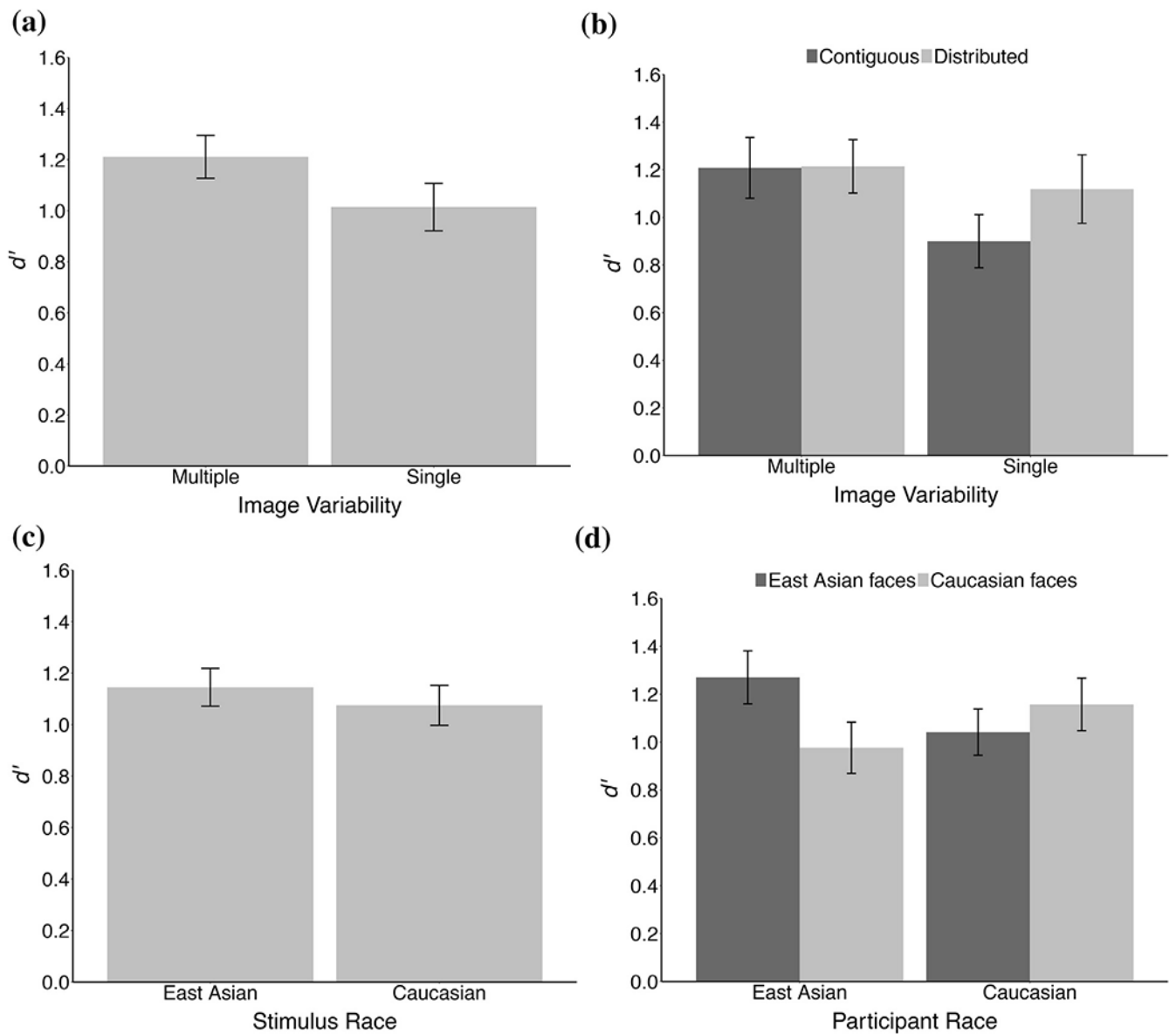
Fig. 8.
Effect of image variability on face recognition accuracy. Results demonstrate greater recognition accuracy for multiple varying images compared to single repeated images (a), but no interaction between learning context and image variability (b). Recognition accuracy was greater for East Asian faces compared to Caucasian faces (c). A robust other-race effect was found (d). Error bars show 95% CI.

Table 1

Participant birth nation for Caucasian and East Asian participants.

Demographics	Experiment 1		Experiment 2[†]	
	Caucasian	East Asian	Caucasian	East Asian
Participant Race	74	59	74	64
Born in the USA	62	30	67	44
Born in another country	12	29	7	20
European Nation	8	0	4	0
East Asian Nation	0	28	0	19
Other	4	1	4	1

Note.

[†]One East Asian and one Caucasian participant did not complete the survey.

Table 2

Country of high school education for Caucasian and East Asian participants.

Demographics	Experiment 1		Experiment 2[†]	
	Caucasian	East Asian	Caucasian	East Asian
United States	71	55[*]	71	61
Other country	3	5	3	3

Note.

[*] One participant attended high school in the United States and Vietnam.

[†] One East Asian and one Caucasian participant did not complete the survey.

Table 3

Native language for Caucasian and Asian participants.

Demographics	Experiment 1		Experiment 2[†]	
	Caucasian	**East Asian**	**Caucasian**	**East Asian**
English	60	8	67	8
English and another European language	2	0	4	0
English and another East Asian language	0	24	0	14
European language	9	0	4	0
East Asian Language	0	26	0	36

Note.

[†]One East Asian and one Caucasian participant did not complete the survey.

Table 4

Generational information for Caucasian and East Asian participants.

Demographics	Experiment 1		Experiment 2[†]	
	Caucasian	East Asian	Caucasian	East Asian
Number of parents born in the United States				
Zero	12	54[*]	5	62
One	8	3	13	0
Two	53	1	53	2
Number of grandparents born in the United States				
Zero	14	56	7	63
One	1	0	2	0
Two	10	3	17	0
Three	9	0	1	0
Four	36	0	47	1
Five/Six	1/1	0	0	0

Note.

[*] One East Asian participant did not reply to this question.

[†] One East Asian and one Caucasian participant did not complete the survey.