Paper:
Horry, R., Cheong, W. & Brewer, N. (in press). The other-race effect in perception and recognition: Insights from the complete composite task.. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 508-524.

http://dx.doi.org/10.1037/xhp0000042

The Other-Race Effect in Perception and Recognition: Insights from the Complete Composite Task

Ruth Horry[a12], Winnee Cheong[b], and Neil Brewer[c]

[a]Flinders University, Adelaide, Australia; and Swansea University, Swansea, UK

[b]HELP University, Kuala Lumpur, Malaysia

[c]Flinders University, Adelaide, Australia

Accepted for publication at *Journal of Experimental Psychology: Human Perception and Performance.*

Ruth Horry, Department of Psychology, College of Human and Health Sciences, Vivian Tower, Swansea University, Swansea, SA2 8PP, UK. Email: R.Horry@swansea.ac.uk.

Phone: +44 (0)1792 295927.

Winnee Cheong, Department of Psychology, HELP University, Kuala Lumpur, Malaysia.

Email: cheongwn@help.edu.my

Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, 5001, Australia. Email: Neil.Brewer@flinders.edu.au

[1]Corresponding author.

Abstract

People are more accurate at recognizing faces of their own race than faces from other races, a phenomenon known as the other-race effect. Other-race effects have also been reported in some perceptual tasks. Across three experiments, White and Chinese participants completed recognition tests as well as the complete paradigm of the composite task, which measures participants' abilities to selectively attend to the target region of a face while ignoring the task-irrelevant region of the face. Each task was completed with both own- and other-race faces. At a group level, participants showed significant own-race effects in recognition, but not in the composite task. At an individual difference level, the results provided no support for the hypothesis that a deficit in holistic processing for other-race faces drives the other-race effect in recognition. We therefore conclude that the other-race effect in recognition is not driven by the processes that underpin the composite effect.

Keywords: Face perception, holistic processing, composite task, other-race effect, own-race bias.

The Other-Race Effect in Perception and Recognition: Insights from the Complete Composite Task

The remarkable capacity of the human visual system to discriminate between faces is thought to be underpinned by holistic processing (Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993). Yet people have difficulty discriminating between faces from ethnic and racial groups other than their own (e.g., Herrmann, Schreppel, Jäger, Ehlis, & Fallgatter, 2007; Lindsay, Jack, & Christian, 1991; Walker & Tanaka, 2003), an effect that translates into a robust memory deficit for other-race faces (Malpass & Kravitz, 1969; Meissner & Brigham, 2001; Sporer, 2001). It has been suggested that other-race faces may be processed less holistically than own-race faces (Rossion & Michel, 2011), yet support for this hypothesis varies across tasks (see Hayward, Crookes, & Rhodes, 2013, for a review). Here, we investigated the other-race effect in the complete paradigm of the composite task, which measures failures to selectively attend to one region of a face (Young, Hellawell, & Hay, 1987; Richler, Cheung, & Gauthier, 2011a). Across three experiments, we adopt an individual differences approach, correlating performance on the composite and recognition tasks. To foreshadow our key findings, the results suggested that other-race faces are processed holistically, and they provided no support for the hypothesis that a holistic processing impairment drives the other-race effect in recognition.

The Other-Race Effect

The other-race effect is the widely replicated finding that people generally recognize faces of their own race more accurately than faces of other races (Malpass & Kravitz, 1969; Meissner & Brigham, 2001). Two main theoretical explanations of the other-race effect in recognition have been advanced (for a recent review, see Hugenberg, Young, Bernstein, & Sacco, 2010). One account suggests that extensive experience with own-race faces optimizes the perceptual system for individuating own-race faces, but that this experience does not generalize well to other-race faces. Recent studies with infants have provided compelling evidence that the other-race effect emerges over the first nine months of life (Kelly et al., 2007). However, with sufficient exposure to other-race faces during this period, the emergence of the other-race effect can be delayed (Heron-Delaney et al., 2011; Sangrigoli & de Schonen, 2004; see also Spangler et al., 2013). The other account suggests that social categorization processes influence the way that other-race faces are perceived and recognized because out-group categorization of other-race faces is automatic (Levin, 2000), and leads to fewer cognitive resources being expended on individuation (Hugenberg et al., 2010; Sporer, 2001). Compelling support for this account comes from studies that have found recognition biases analogous to the other-race effect for ambiguous-race faces (e.g., MacLin & Malpass, 2001) and for own-race faces that are considered to be out-group members along some other salient social dimension (Bernstein, Young, & Hugenberg, 2007; Rule, Ambady, Adams, & MacRae, 2007). Given the strong support for the roles of experience and categorization in the other-race effect, recent models have attempted to synthesize the two positions (Hugenberg, Wilson, See, & Young, 2013).

Though the other-race effect has been investigated most extensively in recognition memory, over recent years many studies have examined the other-race effect in perception using a wide variety of methods (e.g., Golby, Gabrieli, Chiao, & Eberhardt, 2001; Goldinger, He, & Papesh, 2009; Herrmann et al., 2007). This body of research strongly suggests that the other-race effect in recognition stems, at least in part, from how faces are encoded. But what mechanism could underpin differences in encoding? The most thoroughly explored possibility is that other-race faces engage holistic processing to a lesser degree than own-race faces, either through lack of expertise (Bukach, Cottle, Ubiwa, & Miller, 2012) or reduced motivation to individuate (Michel, Corneille, & Rossion, 2007). Holistic processing is widely considered to be the hallmark of human face perception, allowing us to discriminate between many thousands of faces with relative ease (see Maurer et al., 2002, for a review), though researchers debate whether holistic processing is specific to faces or whether it is co-opted for objects of high visual expertise (Robbins & McKone, 2007, *versus* A. C.-N. Wong, Palmeri, & Gauthier, 2009).

Methodologies for Studying Holistic Processing

Two methodologies have been widely used for studying holistic processing: The part-whole task (Tanaka & Farah, 1993), and the same/different composite task (Hole, 1994; hereafter, the composite task). In the part-whole task, participants attempt to recognize facial features shown in isolation or in the context of a whole face. Holistic processing is inferred from a recognition advantage for whole-face trials over part trials. Using the part-whole task, several studies have documented a larger whole-part advantage for own-race faces than for other-race faces in White participants, but (for as yet unknown reasons) not in Asian participants (Crookes, Favelle, & Hayward, 2013; DeGutis, Mercado, Wilmer, & Rosenblatt, 2013; Michel, Caldara, & Rossion, 2006; Mondloch et al., 2010; Tanaka, Kiefer, & Bukach, 2004).

In the composite task, two face halves are combined to create composite faces. Participants study one composite face and are then tested with another. They are required to make rapid same/different judgments about one half of the face (e.g., the top half) while ignoring the other half (e.g., the bottom half). Usually, two types of trials are compared: trials in which the two face halves are horizontally aligned, thus creating the percept of a whole face; and trials in which the two face halves are horizontally misaligned, disrupting the percept of a whole face (e.g., Michel, Rossion, Han, Chung, & Caldara, 2006; Richler, Cheung, et al., 2011a). Holistic processing is inferred if the task-irrelevant part of the face interferes with performance (by reducing accuracy and/or increasing reaction times) when the face halves are aligned to a greater degree than when they are misaligned. Thus, the composite task provides a measure of participants' abilities to selectively attend to one region of a face while ignoring another region of the face.

There are two versions of the composite task – the "partial" and "complete" paradigms (see Figure 1). In the partial paradigm, the top half of the test face can be the same as or different to the studied face, but the bottom half is *always different* to the studied face (e.g., Hole, 1994; Michel, Rossion, et al., 2006; Weston & Perfect, 2005). In the complete paradigm, both the top half and the bottom half can be the same as or different to the studied face, creating the four types of trial shown in Figure 1 (e.g., Farah, Wilson, Drain, & Tanaka, 1998; Gauthier, Curran, Curby, & Collins, 2003; Richler, Cheung, et al., 2011a). Congruent trials (unshaded in Figure 1) are those in which the two face halves are both the same as, or both different to the studied face. These trials do not require selective attention, as a participant could make a fast and accurate response by attending to the whole face. Incongruent trials (shaded in grey in Figure 1) are those in which one half of the test face is the same as the studied face but the other half is different to the studied face. Selective attention to the task-relevant part of the face is required to respond accurately in these trials. Thus, holistic processing is inferred from the difference in performance between congruent and incongruent trials.

In the complete paradigm, discriminability is calculated from the hit rate ("same" responses to same trials) and the false alarm rate ("same" responses to different trials). Holistic processing is inferred if discriminability is higher on congruent trials than on incongruent trials (e.g., Gauthier et al., 2003). Furthermore, this congruency effect is larger when the face halves are aligned than when they are misaligned, creating a Congruency × Alignment interaction (e.g., DeGutis, Wilmer, Mercado, & Cohan, 2013; Z. Gao, Flevaris, Robertson, & Bentin, 2011; Richler, Cheung, et al., 2011a; Richler, Cheung, & Gauthier, 2011b). In the partial paradigm, only same trials are analysed. Holistic processing is inferred from an alignment effect, whereby accuracy is higher (or reaction times are slower) on misaligned trials than on aligned trials (e.g., Hole, 1994; Konar, Bennett, & Sekuler, 2010; Michel et al., 2007; Michel, Rossion, et al., 2006). Thus, the partial paradigm looks only at the hit rate, which leaves it vulnerable to response biases (see Richler, Cheung, et al., 2011b, for a detailed critique of the partial paradigm). Response bias has been shown to vary with congruency (Z. Gao et al., 2011; Richler, Cheung et al., 2011b), alignment (Cheung, Richler, Palmeri, & Gauthier, 2008; Richler, Bukach, & Gauthier, 2009), and their interaction (Z. Gao et al., 2011; Richler, Gauthier, Wenger, & Palmeri, 2008). Crucially, there is no way to control for these response biases in the partial paradigm, as congruency is fully confounded with the correct response (same trials are always incongruent, while different trials are always congruent – see Figure 1).

The complete and partial paradigms have produced qualitatively different results on a number of issues, including the role of low spatial frequencies in face perception (Cheung et al., 2008, *vs.* Goffaux & Rossion, 2006) and the influence of local and global priming on face perception (Z. Gao et al., 2011, *vs*. Weston & Perfect, 2005). It is, therefore, important to consider which paradigm provides the better construct validity. As already argued, the partial paradigm is susceptible to response biases that are unrelated to holistic processing (Richler, Cheung, et al., 2011b). Furthermore, the complete paradigm correlates with another common measure of holistic processing, the part-whole task, whereas the partial paradigm does not (DeGutis, Wilmer, et al., 2013; Wang, Li, Fang, Tian, & Liu, 2012). Thus, the complete paradigm has convergent validity that is lacking for the partial paradigm. And importantly, individual differences in holistic processing as measured by the complete paradigm have been found to correlate with individual differences in face recognition (DeGutis, Wilmer, et al., 2013; Richler, Cheung, et al., 2011a), whereas individual differences in holistic processing as measured by the partial paradigm have not (Konar et al., 2010, Wang et al., 2012).

It is important to note that, from the participant's perspective, the partial and complete paradigms are identical. In both cases, participants make same/different judgments about one part of a stimulus while attempting to ignore an irrelevant part. There is no theoretical reason to expect participants to use different strategies or to recruit different perceptual mechanisms across the two tasks. Indeed, performance on the partial paradigm trials is similar whether they are incorporated into the complete paradigm or run in isolation (Richler, Cheung, et al., 2011a). Thus, qualitative differences in the results arising from the partial and complete paradigms cannot be explained by different perceptual or cognitive mechanisms. Rather, these differences seem to be a consequence of the poor construct validity of the partial paradigm (Richler & Gauthier, 2014).

A number of studies have investigated the other-race effect (and other in-group/out-group effects) using the composite task. In fact, the other-race effect is a prime example of how the partial and complete paradigms have led to qualitatively different conclusions. Of four studies using the partial paradigm, three reported a larger composite effect for own-race (or in-group) faces than for other-race (or out-group) faces (Hugenberg & Corneille, 2009; Michel, Rossion, et al., 2006; Michel et al., 2007), while one reported no significant difference between own- and other-race faces (Mondloch et al., 2010). In contrast, at least to the authors' knowledge, three published papers using the complete paradigm have found that the composite effect is similarly strong for other-race faces as for own-race faces (Bukach et al., 2012; Curby, Johnson, & Tyson, 2012; Harrison, Gauthier, Hayward, & Richler, 2014). Given that there is now compelling evidence to favour the complete paradigm over the partial paradigm (Richler & Gauthier, 2014), the extant data suggest that, at a group level, there is no difference in the magnitude of the composite effect for own- and other-race faces.

In this paper, we take an individual differences approach to other-race face perception. Across three experiments, White and Chinese participants completed composite tasks (using the complete paradigm) and recognition tasks for own- and other-race faces. We first correlated recognition accuracy with the composite effect, separately for own- and other-race faces. Two recent studies have shown that individual differences in the magnitude of the composite effect positively correlate with face recognition ability (DeGutis et al, 2013; Richler, Cheung, et al., 2011a), but to our knowledge, this correlation has not yet been examined for other-race faces. We then correlated own- and other-race performance within each task. If participants recruit similar perceptual mechanisms for processing own- and other-race faces, then we would expect to find positive correlations between the own-race composite effect and the other-race composite effect. Such a finding would parallel that found by DeGutis, Mercado, et al. (2013) in the context of the part-whole task. Finally, we correlated the magnitude of the “other-race effects” in recognition and in the composite task. Though prior research led us to expect that we would find no group-level other-race effect in the composite task (Bukach et al., 2012; Curby et al., 2012; Harrison et al., 2014), this does not necessarily preclude the possibility that individuals who show a strong own-race effect in recognition will also show an other-race effect in the composite task. Such a correlation has, indeed, been found for the part-whole task (DeGutis, Mercado, et al., 2013). Furthermore, individual differences in contact with other-race individuals have also been found to correlate with the other-race effect in the composite task (Bukach et al., 2012).

In summary, across three experiments, we examined individual differences in the composite task for own- and other-race faces, and we explored the relationship between the composite effect and recognition. Based on prior research, we expected to find that: i) the composite effect would be positively correlated with recognition accuracy, for both own- and other-race faces; ii) the composite effect for own-race faces would be correlated with the composite for other-race faces; and iii) the other-race effect in recognition would be correlated with the other-race effect in the composite task.

Experiments 1 and 2

Experiments 1 and 2 followed the same procedure, and so their methods and results are described together. The only difference between the two experiments was in the stimulus materials used. In both experiments, participants completed the full design of the composite task and an old/new recognition test for White and Chinese faces.

Method

Participants and Design. In both experiments, two groups of participants were recruited, from Australia (White participants) and Malaysia (Chinese participants). In Experiment 1, the Australian participants were 69 undergraduates, postgraduates, and staff at Flinders University. Sixty-five participants who self-identified as White were retained in the sample. Two participants were excluded for using the incorrect response keys, and two were excluded for chance performance on the composite task. Of the remaining 61 participants, 41 (69.5%) were female. The mean age was 25.6 years (range 18 to 52 years). The Chinese participants in Experiment 1 were 58 undergraduates from HELP University (Kuala Lumpur, Malaysia). Twenty participants were excluded because they did not provide any information on their ethnic origin, one was excluded for using the incorrect response keys, and one was excluded for chance performance on the composite task. Of the remaining 36 participants, 27 (75.0%) were female. The mean age was 20.1 years (range 18 to 24 years). All identified as being of Chinese descent.

In Experiment 2, the Australian participants were 65 undergraduates, postgraduates, and staff at Flinders University. Fifty-six participants who self-identified as White were retained in the sample, of whom 34 (60.7%) were female. The mean age was 24.9 years (range 18 to 56 years). The Chinese participants in Experiment 2 were 69 undergraduates from HELP University (Kuala Lumpur, Malaysia). Sixty-six participants who self-identified as being of Chinese descent were retained in the sample, of whom 45 (68.2%) were female. The mean age was 20.5 years (range 18 to 26 years).

The composite task followed a 2 (Participant Race: Chinese vs. White) × 2 (Face Race: own-race vs. other-race) × 2 (Congruency: congruent vs. incongruent) × 2 (Alignment: aligned vs. misaligned) mixed design, with repeated measures on the last three factors. The dependent variable was discriminability (*d′*) (response bias and correct response times were also analyzed; see Supplemental Material for details). The recognition task followed a 2 (Participant Race: Chinese vs. White) × 2 (Face Race: own-race vs. other-race) mixed design, with repeated measures on the second factor. Again, the dependent variable was discriminability (*d′*).

Materials. In Experiment 1, the composite task stimuli were 20 Chinese faces (10 male) from the CAS-PEAL database (W. Gao et al., 2008) and 20 White faces (10 male) from a database provided by R. C. L. Lindsay (Queen’s University, Canada). In Experiment 2, the composite task stimuli were 20 male Chinese faces from the Oriental Face Database (Artificial Intelligence and Robotics, Xi’an Jiaotong University, China) and 20 male White faces from the PUT Face Database (Kasiński, Florek, & Schmidt, 2008).

The faces were cropped to fit inside a standardized oval shape, 168 × 224 pixels in size. None had distinguishing marks, facial hair, spectacles, or visible hairline. The images were converted to grayscale, and were approximately matched for luminance. Each image was bisected approximately half way along the length of the nose (112 pixels high). The top and bottom halves of each image were separated by a 3-pixel gap (Richler, Cheung, et al., 2011a). The stimuli were checked to make sure that no valid combinations would result in a grotesque or implausible appearance because of feature spacing. For the practice trials, 10 photographs of houses were cropped to a standard size, and were bisected along the midline.

For the recognition task, 40 Chinese faces (20 male) and 40 White faces (20 female) were selected from the same databases as those used in the composite task (no individuals appeared in both tasks). Two photographs of each person were used, one for study and one for test, to reduce the possibility that participants would respond to low-level features of the image rather than to the face itself. In Experiment 1, the two images showed the individual with different facial expressions (neutral at study, smiling at test); in Experiment 2, the two images showed different viewpoints (frontal view at study, three-quarter profile at test). Within each ethnic group, faces were randomly split into two sets (each containing 10 males and 10 females). The assignment of these sets to targets and distractors was counterbalanced across participants.

For Experiment 1, the faces were converted to grayscale and were cropped to fit inside an oval, removing the external contour of the face and much of the hairline. To improve accuracy in Experiment 2 (see Ellis, Shepherd, & Davies, 1979; Sporer & Horry, 2011), the faces were shown in full color, and the external features and hairline were retained. Background information and clothing cues were removed in Adobe Photoshop.

Procedure. Experiments 1 and 2 followed an identical procedure. Participants signed up for a study on perceiving and remembering faces. They were told that there would be two different tasks – a face perception task and a face memory task. Within each task, there were two separate blocks – one for own-race faces and one for other-race faces. The orders of the tasks, and the order of the blocks within each task, were counterbalanced across participants.

The procedure for the composite task closely followed that reported by Richler, Cheung, et al. (2011a). Each trial began with a study composite, consisting of randomly combined top and bottom face halves of the same race and gender. To provide a strong test of holistic processing, the top and bottom halves were separated by a white line (three pixels thick) so that it was unambiguous where one image ended and the other began. All study composites were aligned. The study composite remained on screen for 200 ms. The study composite was followed by a blank screen for 500 ms. The test composite then appeared, either aligned or misaligned, for 200 ms. Finally, participants saw a screen reminding them of their response options. Participants responded by pressing one of two labelled keys (the 'z' and 'm' keys) with their index fingers. The key assignment was counterbalanced across participants. Participants were always responding to the top halves of the faces, and were instructed to ignore the bottom halves of the images. They were also told to respond as quickly and accurately as possible. The next trial began 500 ms after the participant's response. All participants began with a practice block, consisting of 16 trials with houses as stimuli. The experimental blocks each included 80 trials (10 of each combination of same/different, congruent/incongruent, and aligned/misaligned).

For the recognition task, participants studied 20 faces (10 female), each shown for 5000 ms with a 1000 ms inter-stimulus interval. Immediately following the study list, participants saw the test list consisting of 40 faces (20 female), half of which were old. For each trial, the participant was asked whether the face was old or new, and was then asked to provide a confidence rating from 50% to 100% in 10% increments. Participants completed two blocks (one for own-race faces and one for other-race faces) in a counterbalanced order.

Results and Discussion

Signal detection estimates of discriminability (*d'*) were calculated (see Green & Swets, 1966; Macmillan & Creelman, 1991) for the composite and recognition tasks. A *d'* value of 0 indicates that the participant was unable to discriminate between same and different trials (or old and new trials). Higher, positive values of *d'* indicate better discrimination. Analyses of response bias and reaction time can be found in the Supplemental Materials. Three measures of effect size are used. For interaction terms, we report partial eta-squared (η_p^2). For repeated measures comparisons, we report the standardized mean change (*d*) with 95% Confidence Intervals (CIs) (Viechtbauer, 2007). For between-groups comparisons, we report Cohen's *d* with 95% CIs (Smithson, 2003).

Composite task. Discriminability in the complete composite task is shown in Figures 2 (Experiment 1) and 3 (Experiment 2). The main effects for face race, congruency, and alignment were significant in both experiments. Discriminability was higher for own-race faces than other-race faces (Experiment 1: $F(1, 95) = 8.54$, $p = .004$, $d = 0.35$, 95% CI [0.14, 0.55]; Experiment 2, $F(1, 120) = 8.85$, $p = .004$, $d = 0.28$ [0.09, 0.45]), for congruent trials than incongruent trials (Experiment 1: $F(1, 95) = 117.89$, $p < .001$, $d = 1.15$ [0.89, 1.40]; Experiment 2: $F(1, 120) = 97.69$, $p < .001$, $d = 0.89$ [0.68, 1.10]), and for aligned trials than misaligned trials (Experiment 1: $F(1, 95) = 39.16$, $p < .001$, $d = 0.71$, [0.49, 0.93]; Experiment 2: $F(1, 120) = 71.76$, $p < .001$, $d = 0.77$ [0.57, 0.97]).

In both experiments, the two-way interaction between Congruency and Alignment was significant: Experiment 1, $F(1, 95) = 80.85$, $p < .001$, $\eta_p^2 = .46$; Experiment 2, $F(1, 120) = 126.45$, $p < .001$, $\eta_p^2 = .51$. As can be seen in Figures 2 and 3, the congruency effect was larger on aligned trials (Experiment 1: $t(96) = 13.23$, $p < .001$, $d = 1.34$ [1.07, 1.62]; Experiment 2: $t(121) = 12.56$, $p < .001$, $d = 1.13$ [0.90, 1.36]) than on misaligned trials (Experiment 1: $t(96) = 3.63$, $p < .001$, $d = 0.37$ [0.16, 0.58]; Experiment 2: $t(121) = 4.63$, $p < .001$, $d = 0.42$ [0.23, 0.60]). This interaction is the hallmark composite effect within the complete paradigm of the task (Richler, Cheung, et al., 2011a).

Crucially, in neither experiment was the Congruency × Alignment × Face Race interaction significant; Experiment 1: $F(1, 95) = 0.14$, $p = .71$, $\eta_p^2 < .01$; Experiment 2: $F(1, 120) = 0.45$, $p = .50$, $\eta_p^2 < .01$. The effect sizes for these interaction terms were close to zero, indicating that the size of the composite effect (the Congruency × Alignment interaction) was similar for own- and other-race faces.

However, the four-way interaction reached significance in Experiment 2, $F(1, 120) = 5.35$, $p = .02$, $\eta_p^2 = .04$ (though not in Experiment 1: $F(1, 95) = 0.01$, $p = .99$, $\eta_p^2 < .01$). To interpret this interaction, the data for Experiment 2 were split by participant race. For the White participants, the Face Race × Congruency × Alignment interaction was significant, $F(1, 55) = 5.28$, $p = .03$, $\eta_p^2 = .09$. Contrary to previous research, the effect size for the Congruency × Alignment interaction was larger for other-race faces, $F(1, 55) = 48.50$, $p < .001$, $\eta_p^2 = .47$, than for own-race faces, $F(1, 55) = 19.21$, $p < .001$, $\eta_p^2 = .26$. For the Chinese participants, however, the three-way interaction did not reach significance, $F(1, 65) = 1.23$, $p = .27$, $\eta_p^2 = .02$, indicating similar composite effects for own-race faces and other-race faces.

In Experiment 1, one additional interaction term reached significance: Alignment × Participant Race: $F(1, 95) = 16.19$, $p < .001$, $\eta_p^2 = .15$. Post-hoc tests showed a significant alignment effect for the White participants, $t(60) = 6.06$, $p < .001$, $d = 0.99$ [0.69, 1.30], but not for the Chinese participants, $t(35) = 1.54$, $p = .13$, $d = 0.27$ [-0.06, 0.61]. No other effects were significant in either experiment.

The results of the composite task were in line with several recent studies that have reported strong composite effects for other-race faces (Bukach et al., 2012; Curby et al., 2012; Harrison et al., 2014). There was no evidence that the magnitude of the composite effect varied for own- and other-race faces and, with effect sizes for the critical interactions close to zero, it is unlikely that we simply lacked the statistical power to detect real, but small, differences. These results add weight to the growing body of literature that suggest that other-race faces do, indeed, engage holistic processing.

Recognition task. Discriminability (d') was analyzed in a 2-way (Face Race × Participant Race) mixed ANOVA. Descriptives for Experiments 1 and 2 are shown in Table 1. Overall, discriminability was higher for own-race than other-race faces in both experiments: Experiment 1, $F(1, 95) = 11.81$, $p = .001$, $d = 0.41$ [0.20, 0.61]; Experiment 2, $F(1, 120) = 36.63$, $p < .001$, $d = 0.47$ [0.28, 0.66]. However, the Participant Race × Face Race interaction was significant in both cases: Experiment 1, $F(1, 95) = 4.47$, $p = .04$, $\eta_p^2 = .05$; Experiment 2: $F(1, 120) = 11.78$, $p = .01$, $\eta_p^2 = .09$. Whereas the White participants showed a significant other-race effect (Experiment 1: $t(60) = 4.52$, $p < .001$, $d = 0.57$ [0.30, 0.84]; Experiment 2: $t(55) = 5.74$, $p < .001$, $d = 0.76$ [0.46, 1.05]), the Chinese participants did not (Experiment 1: $t(35) = 0.84$, $p = .40$, $d = 0.14$ [-0.19, 0.47], Experiment 2: $t(65) = 1.82$, $p = .07$, $d = 0.22$ [-0.02, 0.46]).

Individual differences analyses. Our first analysis explored correlations between the composite effect and recognition accuracy for both own- and other-race faces. To estimate each participant's "composite effect", we subtracted d' (incongruent, aligned) from d' (congruent, aligned) separately for own- and other-race faces.[1]

A necessary step in correlational analyses (though one that is often overlooked) is to determine the internal reliabilities of each of the parameter estimates (DeGutis, Wilmer, et al., 2013). These reliability estimates are crucial because they determine the true upper bounds of the correlations, and so are necessary for evaluating the size of a correlation. Full details of the calculations are detailed in the Supplementary Materials, and internal reliability estimates are shown in Tables S1 (Experiment 1) and S2 (Experiment 2). Each correlation plot shows the observed coefficient (r_{observed}), the theoretical upper-bound of the correlation, and the adjusted coefficient (r_{adjusted}), which is simply r_{observed} / upper-bound. Prior to each analysis, outliers (participants more than 2.5 standard deviations from the mean on either variable) were removed. Consequently, the degrees of freedom differ slightly across coefficients. Within each experiment, 12 correlations were explored; we therefore adopted a Bonferroni-corrected α of .004 (.05/12), two-tailed, for all correlations.

Prior research with the complete composite task has shown that individual differences in the magnitude of the composite effect correlate with individual differences in face recognition (e.g., DeGutis, Wilmer, et al., 2013; Richler, Cheung, et al., 2011a). However, these studies used own-race faces only; to our knowledge, no study has yet explored the relationship between the composite effect and recognition with other-race faces. If other-race recognition is less reliant on holistic processing than own-race recognition (e.g., Michel, Rossion, et al., 2006), we would expect to see a weaker composite-recognition relationship for other-race faces than for own-race faces. However, as shown in Figure 4, all of the correlations were weak and non-significant (maximum r (57) = .24, p = .07).

Our second analysis explored the correlation between own- and other-race performance within each task (see Figure 5). If participants recruit similar perceptual mechanisms for processing own- and other-race faces, we would expect to see a positive relationship between the magnitude of the composite effect for own- and other-race faces. As shown in panels A and C of Figure 5, the coefficients for the Chinese participants were positive and small to moderate in size; however, they did not reach our required α level of .004 (in Experiment 1: r (34) = .35, p = .04; in Experiment 2: r (61) = .31, p = .01). The coefficients for the White participants small and non-significant (in Experiment 1: r (57) = .13, p = .33; in Experiment 2: r(53) =-.03; p = .83).

We also explored the relationship between recognition of own-race faces and recognition of other-race faces (see panels B and D of Figure 5). None of the coefficients approached statistical significance (maximum r (57) = .20, p = .13).

Finally, we explored the relationships between the other-race effects in recognition and the composite task. To examine the relationship between the other-race effects in recognition and in the composite task requires a value for each participant that reflects their 'other-race effect' on each task. Following DeGutis, Mercado, et al. (2013), we used a regression approach to calculate two independent "other-race effect" parameters for each task: an own-race advantage parameter (the degree to which one's performance with own-race faces differs from what would be predicted from one's performance with other-race faces), and an other-race deficit parameter (the degree to which one's performance with other-race faces differs from what would be predicted from one's performance with own-race faces).

For each experiment, we correlated the own-race advantages on the two tasks, and we correlated the other-race deficits on the two tasks (see Figure 6). From prior research (Buckach et al., 2012; DeGutis, Mercado, et al., 2013), we expected these relationships to be significant and positive. However, none of the correlations were statistically significant (maximum r (52) = -.24, p = .08), and in fact, the strongest coefficient was negative (see Panel D).

Regarding the individual differences analyses, we had put forward three hypotheses: 1) The magnitude of the composite effect would be positively correlated with recognition accuracy, for both own- and other-race faces; 2) The composite effect for own-race faces would be positively associated with the composite effect for other-race faces. 3) The magnitude of the other-race effect in recognition would be positively associated with the other-race effect in the composite task. None of these hypotheses were supported by the analyses. Of 24 correlation coefficients from Experiments 1 and 2, none were statistically significant at our Bonferroni-corrected α level of .004. However, these null coefficients must be interpreted very cautiously. A major limitation of these experiments was in the internal reliability of both the recognition and composite tasks. As a direct consequence, the theoretical upper bounds of the correlations were often very low (ranging from .02 to .55). With such low upper bounds, the chances of detecting real relationships are greatly diminished.

Experiment 3

In Experiment 3, we attempted to address the reliability problem as far as was possible. The reliability of the composite task in Experiments 1 and 2 was within the range reported by Ross, Richler, and Gauthier (in press).[2] We focused, therefore, on improving the reliability of the recognition measure, by using two versions of the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006): the CFMT-Chinese (McKone et al., 2012), and the CFMT-Australian (McKone et al., 2011), both of which have been shown to have high internal reliability with own- and other-race samples. Improvements in the reliability of the recognition test should, in turn, have increased the theoretical upper bounds on the correlations between holistic processing and recognition, decreasing the likelihood of Type II errors.

In addition to the change of recognition tests, we also made some changes to the composite task in Experiment 3. Rossion (2013) argued that there are many procedural differences between the studies that have used the complete paradigm (upon which our procedures for Experiments 1 and 2 were based) and the studies that have used the partial paradigm, differences which could have confounded the debate around which measure is more appropriate. Thus, in Experiment 3, we matched our procedure to that used by Michel, Rossion, et al. (2006) (with the exception of the additional complete paradigm trials). We made the following changes: 1) Exposure duration to the study image was increased from 200 ms to 600 ms, and the test image remained on screen until a response was made; 2) The alignment of the study image matched the alignment of the test image, thus reducing the need for lateral shifts in attention between the study and test images; 3) Own- and other-race trials were randomly intermixed within each block; and 4) Face halves were carefully paired based upon internal feature spacing, rather than randomly combined. Note that it was not our intention to systematically evaluate any of these factors; rather, we attempted to ensure that our results were robust to these minor procedural variations.[3]

Method

Participants and Design. Two groups of participants were recruited: one from Malaysia (*N* = 81) and one from Australia (*N* = 59). The final sample consisted of 77 participants from Malaysia who self-identified as Chinese and 51 participants from Australia who self-identified as White. The proportion of female participants was 75.32% for the Chinese sample and 78.43% for the White sample. The mean age of the Chinese participants was 22.16 years (*SD* = 1.01) and the mean age of the White participants was 22.92 years (*SD* = 7.21).

The composite task followed a 2 (Congruency: congruent vs. incongruent) × 2 (Alignment: aligned vs. misaligned) × 2 (Face Race: own-race vs. other-race) × 2 (Participant Race: Chinese vs. White) mixed design, with repeated measures on all but the last factor. The dependent measure was discriminability (d') (response bias and correct RTs were also analyzed; see Supplemental Materials).

The recognition tests followed a 2 (Face Race: own-race vs. other-race) × 2 (Participant Race: Chinese vs. White) mixed design, with repeated measures on the first factor. The dependent measure was number correct (scored out of 72).

Materials. The stimuli for the composite task consisted of 20 male Chinese faces from the Oriental Face Database and 20 male White faces from the PUT Face Database (Kasiński et al., 2008). Each original face was cropped to fit inside a standardized 168 x 224 pixel oval, converted to grayscale, and bisected across the centre of the nose. Each face was carefully paired with another face on the basis of feature spacing (Rossion, 2013). For example, for Face A, which was paired with Face B, four composites were created: AA (Same-Congruent), AB (Same-Incongruent), BA (Different-Incongruent), and BB (Different-

Congruent). A three-pixel gap separated the top and bottom halves of each composite. An aligned and a misaligned version of each composite were created.

The recognition tests were the CFMT-Australian (McKone et al., 2011) and the CFMT-Chinese (McKone et al., 2012), both of which use the Cambridge Face Memory Test format (Duchaine & Nakayama, 2006). Each test includes grayscale images of 52 male faces (six targets and 46 distractors). For the CFMT-Australian, the targets and 8 of the distractors are from the Australian National University Face Database, all of whom are of Australian or New Zealand descent. The remaining distractors are from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). For the CFMT-Chinese, the faces are all of graduate students at the Chinese Academy of Sciences in Beijing, and all individuals are Han Chinese. Internal reliability has been reported as .88 for the CFMT-Australian (McKone et al., 2011). Internal reliability for the CFMT-Chinese has been reported as .90 for Asian participants and .89 for Australian participants (McKone et al., 2012).

Procedure. Participants signed up for a study on perceiving and remembering faces. Each participant completed the CFMT-Australian, the CFMT-Chinese, and two blocks of the composite task (one upright and one inverted; only the data from the upright block are presented here). The order of the tasks was counterbalanced, as was the order of the upright and inverted blocks of the composite task.

Due to the inclusion of the inverted condition in Experiment 3, we made a small change to the instructions. Specifically, participants were instructed to respond to the *forehead* half of the image, rather than the *top* half as in Experiments 1 and 2. Each trial proceeded as follows: the study image was shown for 600 ms, followed by a blank screen for 300 ms. The test image then appeared, and it remained on screen until the participant responded. In each block of the composite task, participants completed 160 trials, which consisted of 80 own-race and 80 other-race trials, randomly intermixed. Each original study face was seen eight times per block, once per cell of the design. Prior to the beginning of each block, the participants completed 16 practice trials with houses as stimuli.

Briefly, the recognition tasks each included six target faces. The test takes a three-alternative forced-choice format, and increases in difficulty over the course of the 72 trials. For full details of the procedure, see Duchaine and Nakayama (2006). Within each of the recognition tests, the order of trials, and the order of the alternatives within each trial, was fixed.

Results and Discussion

Composite task. Discriminability was analyzed in a 4-way (Congruency × Alignment × Face Race × Participant Race) mixed ANOVA (for analyses of correct RTs and partial paradigm trials, see Supplemental Materials). Mean discriminability is shown in Figure 7. Discriminability was higher on congruent trials (M = 2.33, SD = 0.61) than on incongruent trials (M = 1.86, SD = 0.72), $F(1, 126) = 74.54$, $p < .001$, d = 0.79 [0.59, 0.98]. Discriminability was also higher on misaligned trials (M = 2.33, SD = 0.65) than on aligned trials (M = 2.06, SD = 0.62), though the effect was small, $F(1, 126) = 5.65$, $p = .02$, d = 0.19 [0.01, 0.36]. However, both of these main effects should be interpreted in light of the expected and significant Congruency × Alignment interaction, $F(1, 126) = 132.04$, $p < .001$, $\eta_p^2 = .51$. Post-hoc tests showed that there was a large congruency effect on aligned trials, $t(127) = 13.83$, $p < .001$, d = 1.22 [0.99, 1.44], but no significant congruency effect on misaligned trials, $t(127) = 1.22$, $p = .22$, d = 0.11 [-0.07, 0.28]. The crucial Congruency × Alignment × Face Race interaction was not significant, $F(1, 126) = 0.41$, $p = .52$, $\eta_p^2 = .003$, and the effect size was close to zero. The four-way interaction was also non-significant, F(1,

126) = 0.15, p = .70, η_p^2 = .001. Thus, in line with Experiments 1 and 2, these results suggest that participants processed both own- and other-race faces holistically. There was no indication that the magnitude of the composite effect varied for own- and other-race faces, with the effect size of the interaction, once again, very close to zero (see also Bukach et al., 2012; Curby et al., 2012; Harrison et al., 2014). From a methodological perspective, the similarity of the results of Experiments 1 to 3 suggest that the procedural differences between prior partial and complete paradigm studies have not likely contributed to the discrepant findings in the literature, despite the concerns raised by Rossion (2013).

There were, however, some significant effects of Face Race on discriminability. Overall, participants showed greater discriminability for own-race faces (M = 2.14, SD = 0.67) than for other-race faces (M = 2.06, SD = 0.65), F(1, 126) = 10.99, p = .001, d = 0.14 [-0.03, 0.31], though the CI of the effect size included zero. The Country × Face Race interaction was also significant, F(1, 126) = 52.37, p < .70, η_p^2 = .29. Whereas the Chinese participants showed greater discriminability on other-race trials (M = 2.16, SD = 0.67) than on own-race trials (M = 1.99, SD = 0.64), t(76) = 3.20, p = .002, d = -0.36 [-0.59, -0.13], the White participants showed greater discriminability on own-race trials, (M = 2.35, SD = 0.66) than on other-race trials (M = 1.90, SD = 0.59), t(50) = 6.53, p < .001, d = 0.90 [0.58, 1.23]. No other interactions were significant (maximum F(1, 126) = 2.46, p = .12, η_p^2 = .02).

Recognition task. The CFMT is scored as the number of correct responses; the maximum number of correct responses is 72. CFMT scores were analyzed in a 2 (Test: CFMT-Aus vs. CFMT-Chinese) × 2 (Participant Race: Chinese vs. White) mixed ANOVA. Neither the main effect of test, F(1, 126) = 2.72, p = .10, d = 0.25 [0.07, 0.43], nor the main effect of participant race, F(1, 126) = 0.35, p = .56, d = 0.06 [-0.29, 0.42], were significant. However, the interaction term was significant, F(1, 126) = 131.39, p < .001, η_p^2 = .51. Post-hoc tests showed that the Chinese participants were more accurate on the CFMT-Chinese (*M* = 59.00, *SD* = 8.92) than on the CFMT-Australian (*M* = 50.08, *SD* = 6.51), $t(76) = 10.11$, $p < .001$, $d = 1.14$ [0.85, 1.43], whereas the White participants were more accurate on the CFMT-Australian (*M* = 57.04, *SD* = 8.21) than on the CFMT-Chinese (*M* = 49.70, *SD* = 9.02), $t(50) = 6.61$, $p < .001$, $d = 0.91$ [0.59, 1.24].

Individual differences analyses. Following the same procedures as outlined for Experiments 1 and 2, we conducted a series of correlational analyses for Experiment 3. The internal reliabilities of the relevant conditions are shown in Table S3 in the Supplemental Materials. The reliability of the recognition task was substantially higher in Experiment 3 than in Experiments 1 and 2, which consequently increased the upper bounds many of the correlations, rendering the coefficients more readily interpretable. As in the previous experiments, prior to conducting each correlation, we excluded participants further than 2.5 standard deviations from the mean on either measure. Once again, we adopted a Bonferroni-corrected α level of .004 (two-tailed) to correct for multiple tests.

We first examined correlations between the composite effect and recognition accuracy (see Figure 8). None of the correlations reached the pre-specified α level of .004 (maximum $r(47) = .29$, $p = .04$).

Next, we examined correlations between the composite effects for own- and other-race faces (see Figure 9, panel A). For White participants, the coefficient was positive and moderate, though it did not reach the corrected α level of .004, $r(47) = .38$, $p = .007$. The coefficient for Chinese participants was smaller and did not approach statistical significance, $r(73) = .17$, $p = .14$.

Regarding the recognition task, the correlations between own- and other-race recognition were positive and statistically significant: for Chinese participants, $r(73) = .54$, *p*

$< .001$; for White participants, $r(47) = .41$, $p = .003$. It seems, therefore, that the CFMT format captures general recognition abilities that extend beyond one's own race. In other words, people with an above-average recognition ability for own-race faces will likely also have an above-average recognition ability for other-race faces.

Finally, we correlated individual differences in the magnitude of the other-race effects in the composite task and in the CFMT (see Figure 10). None of the correlations were statistically significant (maximum r (48) = .25, p = .08); for the Chinese participants, the coefficients were close to zero. Thus, these correlations provided no support for the hypothesis that the other-race effect in recognition is driven by holistic processing. However, these null correlations should be interpreted cautiously, for two reasons. First, the upper bounds of the correlations were still low, ranging from .26 to .49. Indeed, the adjusted correlations for the White participants were approaching their upper bound (see Figure 10). Second, we were unable to replicate the finding that individual differences in the magnitude of the composite effect correlate with face recognition abilities for own-race faces (e.g., Richler, Cheung, et al., 2011a; DeGutis, Wilmer, et al., 2013), nor did we find such a correlation for other-race faces. If, within our sample, performance on the composite task was not related to recognition, it should come as no surprise that the "other-race effects" on the two tasks were also unrelated.

General Discussion

Across three experiments, with six independent samples of participants, we investigated the relationship between the composite effect and recognition for own- and other-race faces. We first discuss the group level effects, and the methodological and theoretical implications of our findings. We then discuss the individual level effects, and we consider the theoretical implications of our findings for the other-race effect.

At the group level, the results of the composite task were consistent: participants showed strong composite effects for other-race faces. Our results align with those of three others studies that have used the complete composite paradigm to investigate the other-race effect, none of which have reported evidence for superior holistic processing of other-race faces (Bukach et al., 2012; Curby et al., 2012; Harrison et al., 2014). These findings contrast with those from the partial composite paradigm, in which own-race faces were associated with larger composite effects than other-race faces (Michel, Rossion, et al., 2006; Michel et al., 2007; though see Mondloch et al., 2010). What could be driving such a discrepancy? One possibility is that we lacked statistical power to detect a statistically significant Congruency × Alignment × Face Race interaction. However, this seems unlikely, as the effect sizes for the crucial interactions were close to zero, while our sample sizes were reasonably large for this type of study. A second possibility is that the partial and complete paradigms recruit different perceptual mechanisms, and so are measuring different aspects of holistic processing. However, from the participant's perspective, the partial and complete tasks are virtually identical, and there is no theoretical reason to expect that participants would engage different processes in the two tasks. Furthermore, the partial paradigm trials yield similar results whether they are incorporated into the complete paradigm or whether they are conducted in isolation (Richler, Cheung, et al., 2011a). The most convincing explanation for the discrepancy is that the partial paradigm lacks construct validity due to its susceptibility to confounding response biases (Gauthier & Bukach, 2007; Richler, Cheung, et al., 2011b; Richler & Gauthier, 2014).

It is worth noting that our findings, and those of other composite task studies (Bukach et al., 2012; Curby et al., 2012; Harrison et al., 2014) are qualitatively different from those of studies that have used the part-whole task, in which other-race effects have been found consistently (e.g., Crookes et al., 2013; DeGutis, Mercado, et al., 2013; Tanaka et al., 2004; Michel, Caldara, et al., 2006; Mondloch et al., 2010). Though both the composite task and the part-whole task are commonly referred to as tests of holistic processing, they operationalize this concept quite differently. In the composite task, holistic processing is operationalized as the interference between two face halves. In the part-whole task, holistic processing is operationalized as a facilitative effect of the whole face when making judgments about a specific face part. Though there is some evidence that performance on one task can predict performance on the other (DeGutis, Wilmer, et al., 2013; Wang et al., 2012), it is not clear that the two effects rely on the same underlying mechanisms. Richler, Palmeri, and Gauthier (2012) discuss four distinct mechanisms that could produce holistic processing: the use of global face templates (Young et al., 1987); the representation of the spatial relations between features (Diamond & Carey, 1986); parallel processing of multiple features and their spatial relationships (Fific & Townsend, 2010); and an automatic attentional strategy of processing all parts of the face together (Y. K. Wong & Gauthier, 2010). It is quite possible that the composite task and the part-whole task are, in fact, drawing on distinct mechanisms that involve separate neural pathways. It is also possible that the other-race effect can be observed in some of these pathways but not in others, and that it is the interaction between these pathways that creates the robust recognition deficit that has captivated researchers for 45 years (Malpass & Kravitz, 1969; Meissner & Brigham, 2001). Clearly, a goal for future research will be to establish how the different aspects of holistic processing relate to one another, and through which routes the other-race effect emerges.

Despite showing strong composite effects for other-race faces, the participants showed typical other-race effects in the recognition tasks. As is often the case, the other-race effects in the old/new recognition tests were asymmetrical, with significant effects for the

White participant and non-significant effects for the Chinese participants. However, the reliability of these tasks was quite low, suggesting that they did not provide a reliable measure of the participants' recognition abilities. In Experiment 3, the CFMT format allowed us to detect strong other-race effects in both groups of participants, replicating prior research with Australian and Asian samples (McKone et al., 2012). Furthermore, the internal reliability of the CFMT tasks was much higher (ranging from .77 to .90) than the internal reliability of the old/new recognition tasks (which ranged from .05 to .49), which had significant implications for interpretation of the correlational analyses. The methodological implications of these reliability estimates for researchers wishing to take an individual differences approach to face recognition are clear: wherever possible, the CFMT format should be used in favour of old/new recognition tests.

Taken together, the individual differences analyses of Experiments 1 to 3 provided no support for the holistic processing account of the other-race effect. In none of the experiments did we find statistically significant correlations between other-race effects in the two tasks. In fact, leaving the issue of statistical significance aside, several of the coefficients were negative, and many were close to zero. However, perhaps we should not be surprised that these correlations were weak; after all, we were unable to replicate the finding that the magnitude of the composite effect (for own-race faces) correlates with recognition accuracy (as in Richler, Cheung, et al., 2011a; DeGutis, Wilmer, et al., 2013). Why this should be so is not clear. The procedures of Experiments 1 and 2 were closely matched to that used by Richler, Cheung, et al. (2011a), and our sample sizes were considerably larger. One possibility is that the low reliability of the composite task greatly inflated the likelihood of Type II error by reducing the true upper bound of the correlations. However, we should note that Richler, Floyd, and Gauthier (2014), also found that the composite effect did not correlate significantly with CFMT. Those authors speculated that successful performance on the CFMT may be achievable through part-based processing, which is a claim that deserves empirical attention. In summary, while we must be cautious about drawing firm conclusions from null effects, the totality of evidence from these studies argues against the hypothesis that the other-race effect in recognition is driven by differential holistic processing. Instead, the implication seems to be that both own- and other-race faces strongly engage holistic processing.

Acknowledgements

This research was funded by an ARC Discovery Grant to N. Brewer et al. Thanks to Nathan Weber for his assistance with programming and to Gabrielle Comey and Emily Barnett for their assistance with data collection. We are thankful to Elinor McKone and Jia Liu for sharing the materials for the CFMT-Australian and the CFMT-Chinese, respectively.

References

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*, 706-712. doi: 10.1111/j.1467-9280.2007.01964.x.

Bukach, C. M., Cottle, J., Ubiwa, J., & Miller, J. (2012). Individuation experience predicts other-race effects in holistic processing for both Caucasian and Black participants. *Cognition, 123,* 319-324. doi: 10.1016/j.cognition.2012.02.007.

Burton, M. A., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42,* 268-291. doi: 10.3758/brm.42.1.286.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 1327-1336. doi: 10.1037/a0011752.

Crookes, K., Favelle, S., & Hayward, W. G. (2013). Holistic processing for other-race faces in Chinese participants occurs for upright but not inverted faces. *Frontiers in Psychology*, *4,* Article 29. doi: 10.3389/fpsyg.2013.00029.

Curby, K. M., Johnson, K. J., & Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition & Emotion*, *26*, 93-102. doi: 10.1080/02699931.2011.555752.

DeGutis, J., Mercado, R. J., Wilmer, J., & Rosenblatt, A. (2013). Individual differences in holistic processing predict the own-race advantage in recognition memory. *PLoS ONE, 8,* e58253. doi: 10.1371/journal.pone.0058253.

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition, 126*, 87-100. doi: 10.1016/j.cognition.2012.09.004.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115, 107-117. doi: 10.1037/0096-3445.115.2.107

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*, 576-585. doi: 10.1016/j.neuropsychologia.2005.07.001.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*, 431-439. doi:10.1068/p080431.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review, 105*, 482-498. doi: 10.1037/0033-295X.105.3.482

Fific, M., & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: Identifying mechanisms of secondary-level holism within a categorization paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*, 1290–1313. doi: 10.1037/a0020123

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, B. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, And Cybernetics – Part A: Systems and Humans, 38*, 149-161. doi: 10.1109/TSMCA.2007.909557.

Gao, Z., Flevaris, A.V., Robertson, L. C., & Bentin, S. (2011). Priming global and local processing of composite faces: Revisiting the processing-bias effect on face perception. *Attention, Perception, & Psychophysics, 73*, 1477-1486. doi: 10.3758/s13414-011-0109-7.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*, 322-330. doi: 10.1016/j.cognition.2006.05.003.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*, 428-432. doi: 10.1038/nn1029.

Goffaux, V., & Rossion, B. (2006). Faces are “spatial” – Holistic face perception is supported by low special frequencies. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 1023-1039. doi: 10.1037/0096-1523.32.4.1023.

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

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 1105-1122. doi: 10.1037/a0016548.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley & Sons.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99*, 45-56. doi: 10.1348/000712607X199981.

Harrison, S. A., Gauthier, I., Hayward, W. G., & Richler, J. J. (2014). Other-race effects manifest in overall performance, not qualitative processing style. *Visual Cognition, 22,* 843-864. doi: 10.1080/13506285.2014.918912.

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21,* 1224-1247. doi: 10.1080/13506285.2013.824530.

Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J.W., … Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PLoS ONE, 6,* e19858. doi: 10.1371/journal.pone.0019858.

Herrmann, M. J., Schreppel, T., Jäger, D., Ehlis, A.-C., & Fallgatter, A. J. (2007). The other-race effect for face perception: An event-related potential study. *Journal of Neural Transmission*, *114*, 951-957. doi: 10.1007/s00702-007-0624-9.

Hole, G. J. (1994). Configural factors in the perception of unfamiliar faces. *Perception*, *23*, 65-74. doi:10.1068/p230065.

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science*, *33*, 1173-1181. doi: 10.1111/j.1551-6709.2009.01048.x.

Hugenberg, K., Wilson, J. P., See, P. E., & Young, S. G. (2013). Towards a synthetic model of own group biases in face memory. *Visual Cognition, 21*, 1392-1417. doi: 10.1080/13506285.2013.821429.

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. *Psychological Review, 117,* 1168-1187. doi: 10.1037/a0020463.

Kasiński, A., Florek, A., & Schmidt, A. (2008). The PUT Face Database. *Image Processing & Communications, 13*, 59-64.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science, 18,* 1084-1089. doi: 10.1111/j.1467-9280.2007.02029.x.

Konar, Y., Bennett, P. J., & Sekuler, A.B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*, 38-43. doi: 10.1177/0956797609356508.

Levin, D. T., (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*, 559-574. doi: 10.1037/0096-3445.129.4.559.

Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 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. *Psychology, Public Policy, & Law, 7*, 98-118. doi: 10.1037//1076-8971.7.1.98.

Macmillan, N., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330-334. doi: 10.1037/h0028434.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260. doi: 10.1068/p3339.

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R.B., Rivolta, D., … O'Connor, K. B. (2011). Face race and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test-Australian. *Cognitive Neuropsychology, 28*, 109-146. doi: 10.1080/02643294.2011.616880.

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentinu, C., Pidcock, M., … Pelleg, M. (2012). A robust method of measuring other-race and other-race effects: The Cambridge Face Memory Test format. *PLoS One, 7,* e47956. doi: 10.1371/journal.pone.0047956.

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

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*, 55-73. doi: 10.1080/13506280500158761.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science, 31*, 911-924. doi: 10.1080/03640210701530805.

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17*, 608-615. doi: 10.1111/j.1467-9280.2006.01752.x.

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & 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.

Richler, J. J., Bukach, C. M., & Gauthier, I. (2009). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception, & Psychophysics, 71*, 530-540. doi: 10.3758/APP.71.3.530.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological Science*, *22*, 464-471. doi: 10.1177/0956797611401753.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011b). Beliefs alter holistic face processing... if response bias is not taken into account. *Journal of Vision, 11*, 1-13. doi: 10.1167/0.0.1.

Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic processing. *Journal of Vision, 14*, 1-14. doi: 10.1167/14.11.10

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic processing. *Psychological Bulletin*, *140*, 1281-1302. doi: 10.1037/a0037004

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *34*, 328-342. doi: 10.1037/0278-7393.34.2.328.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research, 51*, 333-342. doi: 10.1016/j.visres.2010.11.014.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology, 3,* Article 553. doi: 10.3389/fpsyg.2012.00553.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*, 34-79. doi: 10.1016/S0010-0277(03)00020-9.

Ross, D. A., Richler, J. J., & Gauthier, I. (in press). Reliability of composite-task measures of holistic face processing. *Behavior Research Methods*. doi: 10.3758/s13428-014-0497-4.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*, 139-253. doi: 10.1080/13506285.2013.772929.

Rossion, B., & Michel, C. (2011). An experience-based holistic account of the other-race face effect. In A. Calder, G. Rhodes, M. Johnson, J. Haxby (Eds.). *Oxford Handbook of Face Perception* (p. 215-244). Oxford: Oxford University Press.

Rule, N. O., Ambady, N., Adams, R. B., & MacRae, C. N. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review, 14*, 687-692. doi: 10.3758/BF03196822.

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

Smithson, M. J. (2003). *Confidence Intervals: Quantitative Applications in the Social Sciences Series, No. 140*. Thousand Oaks, CA: Sage.

Spangler, S. M., Schwarzer, G., Freitag, C., Vierhaus, M., Teubert, M., Fassbender, I., … Keller, H. (2013). The other-race effect in a longitudinal sample of 3-, 6- and 9-month old infants: Evidence of a training effect. *Infancy, 18,* 516-533. doi: 10.1111/j.1532-7078.2012.00137.x

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, & Law, 7*, 36-97. doi: 10.1037//1076-8971.7.1.36.

Sporer, S. L., & Horry, R. (2011). Recognizing faces from ethnic in-groups and out-groups: Importance of outer face features and effects of retention interval. *Applied Cognitive Psychology, 25*, 424-431. doi: 10.1002/acp.1709.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology Section A*, 46, 225-245. doi: 10.1080/14640749308401045.

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.

Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics, 32,* 39-60. doi: 10.3102/1076998606298034.

Walker, P. M. & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, 32, 1117-1125. doi: 10.1068/p5098

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23*, 169-177. doi: 10.1177/0956797611420575.

Weston, N. J., & Perfect, T. J. (2005). Effects of processing bias on the recognition of composite face halves. *Psychonomic Bulletin & Review, 12*, 1038-1042. doi: 10.3758/BF03206440.

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science, 20*, 1108-1117. doi: 10.1111/j.1467-9280.2009.02430.x.

Wong, Y. K., & Gauthier, I. (2010). Holistic processing of musical notation: Dissociating failures of attention in experts and novices. *Cognitive, Affective, & Behavioral Neuroscience, 10,* 541-551. doi: 10.3758/13.2.22

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception, 16*, 747-759. doi:10.1068/p160747.

Footnote

[1]In previous individual differences research (e.g., Richler, Cheung, et al., 2011), participants' "composite effect" scores were calculated as: ([*d′* Congruent Aligned - *d′* Incongruent Aligned] - [*d′* Congruent Misaligned - *d′* Incongruent Misaligned]). However, this measure is a difference of differences, which tends to lead to very poor reliability. Furthermore, Richler & Gauthier (2014) recently demonstrated that very little variance is shared between the aligned and misaligned congruency effects. We therefore used only the aligned congruency effect as a measure of holistic processing. We note that the coefficients changed very little if the misaligned congruency score was subtracted from the aligned congruency score. However, as expected, the reliability of the parameters was greatly reduced.

[2]Since Experiment 3 was conducted, a more reliable version of the composite task has been published for use in individual difference research – the Vanderbilt Holistic Face Processing Test (VHFPT; Richler, Floyd, & Gauthier, 2014). The reliability of the composite effect in the VHFPT was reported as .56, which is somewhat higher than the reliabilities of our tests, which ranged from .27 to .62. However, the VHFPT includes only White faces, and so cannot, in its current form, be used for investigating other-race effects in holistic processing.

[3]We also included an inverted condition in Experiment 3, in which all stimuli were rotated by 180 degrees. In contrast to previous research (e.g., Curby et al., 2013; Richler, Mack, Palmeri, & Gauthier, 2011), we did not find a significant composite effect for inverted faces, nor did we find any differences between the processing of inverted own-race and other-race stimuli. For brevity, we omit the analyses of the inverted trials here. Interested readers should contact the first author for further information.

Table 1

Mean discriminability (d') for the recognition task in Experiments 1 and 2. Standard deviations are shown in parentheses.

	White participants		Chinese participants	
Experiment	Own-race	Other-race	Own-race	Other-race
1	0.82 (0.50)	0.46 (0.38)	0.57 (0.44)	0.49 (0.52)
2	1.11 (0.58)	0.52 (0.48)	0.82 (0.53)	0.67 (0.50)

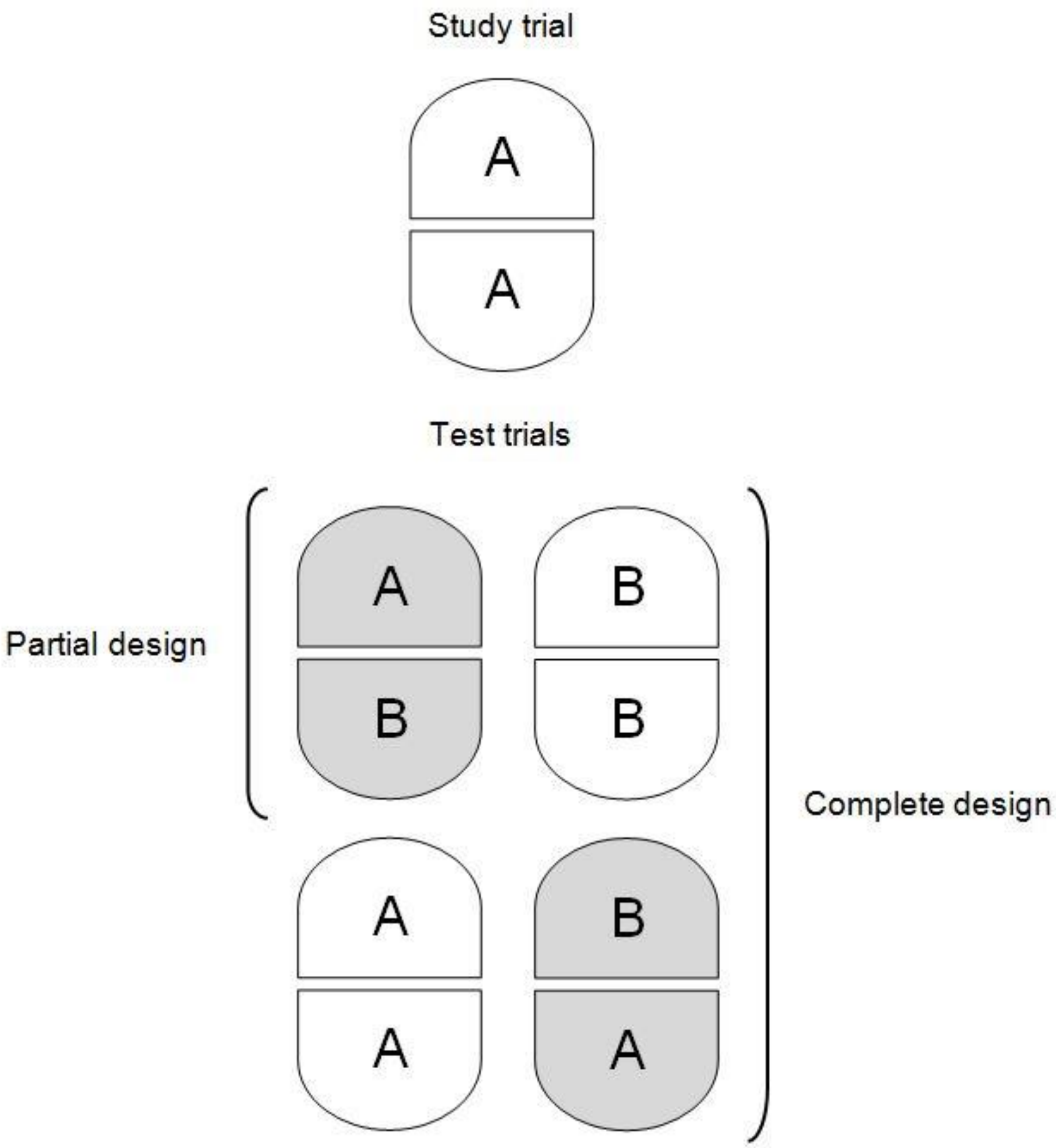

Figure 1. Trial types included in the partial and complete paradigms of the composite task. Incongruent trials are shaded in gray.

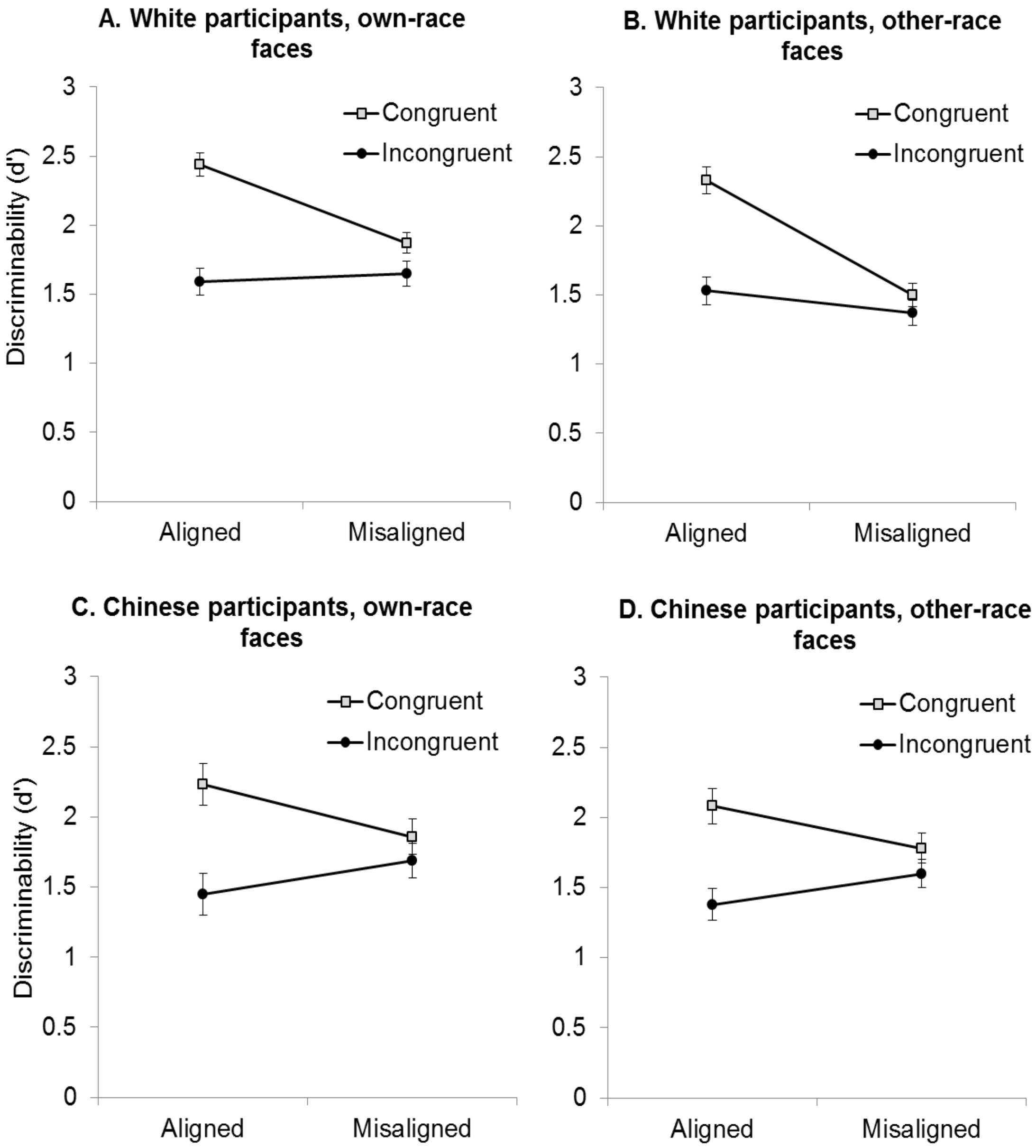

Figure 2. Discriminability (d') in Experiment 1. Error bars represent standard error of the mean.

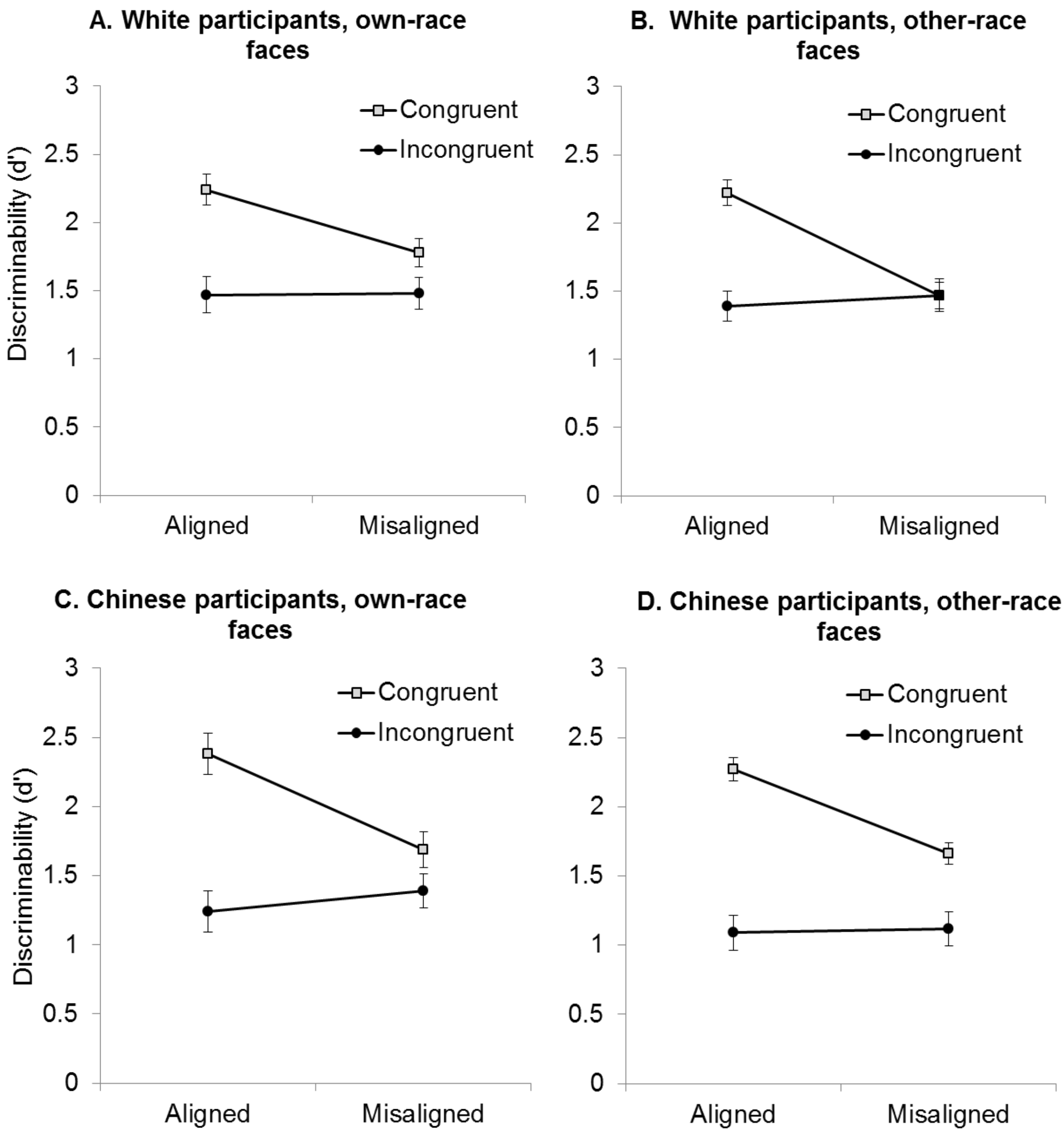

Figure 3. Discriminability (*d′*) in Experiment 2. Error bars represent standard error of the mean.

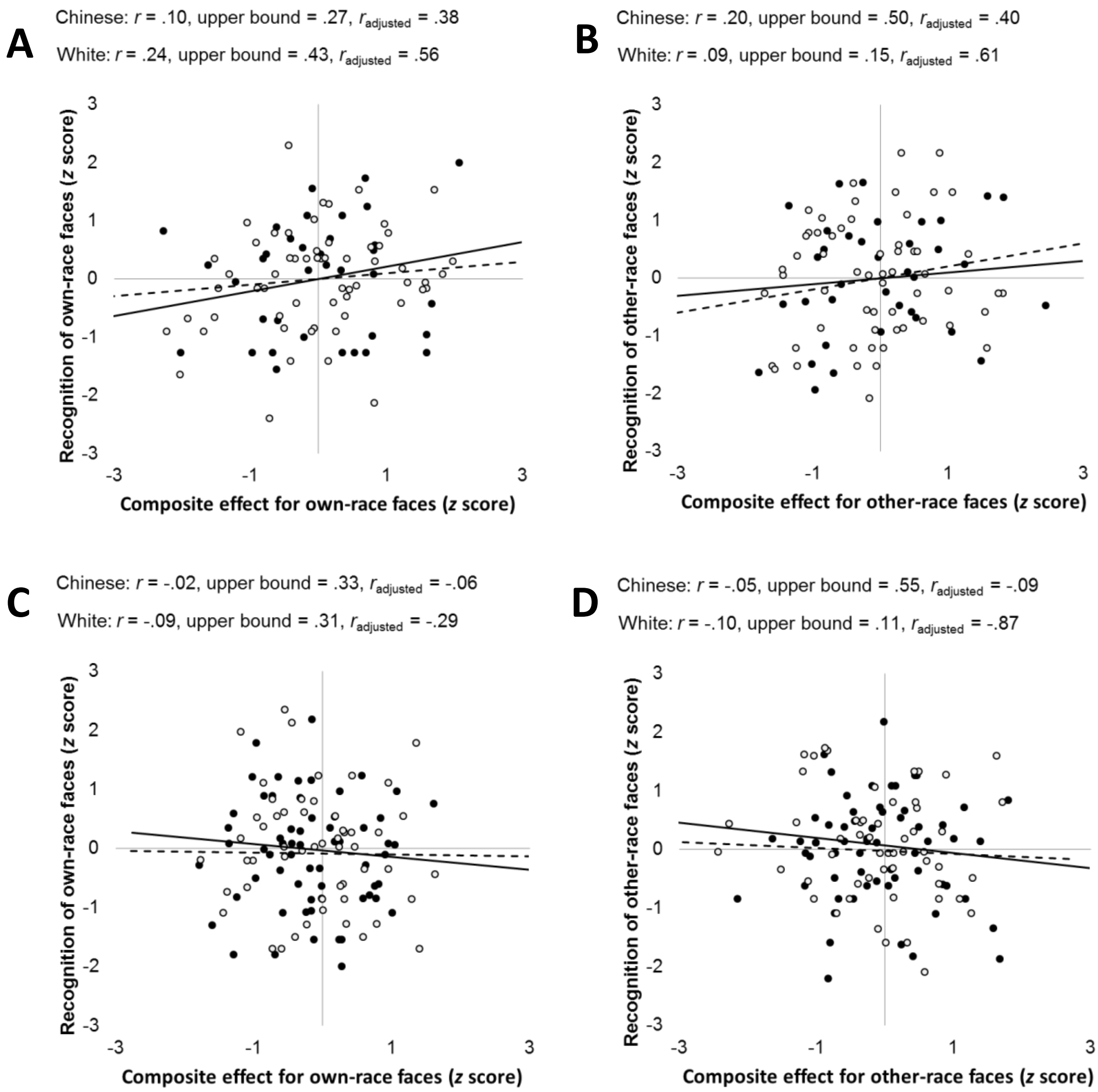

Figure 4. Within-race correlations between holistic processing and recognition in Experiments 1 (A and B) and 2 (C and D). Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.

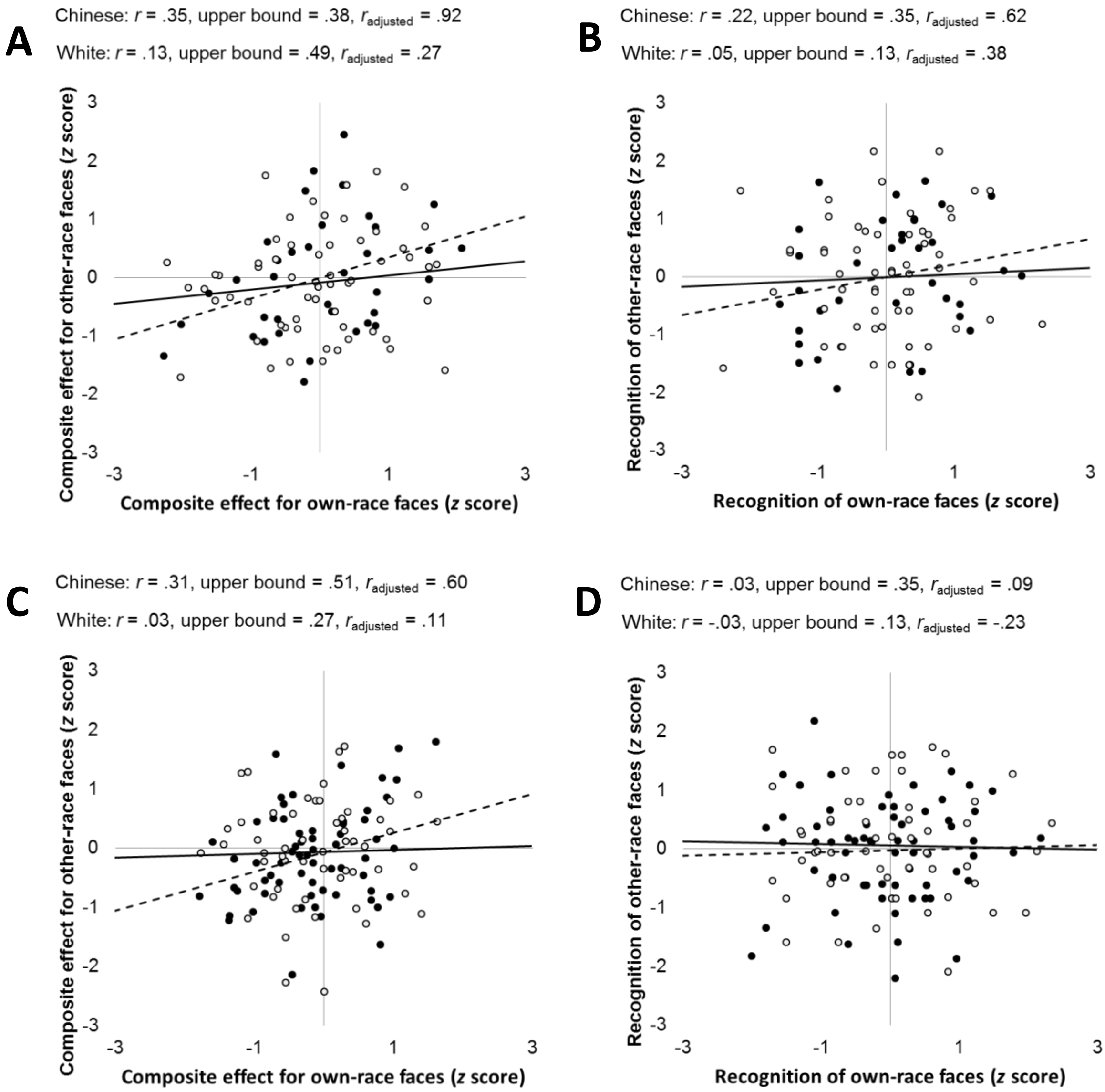

Figure 5. Within-task correlations between own- and other-race performance in Experiments 1 (A and B) and 2 (C and D). Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.

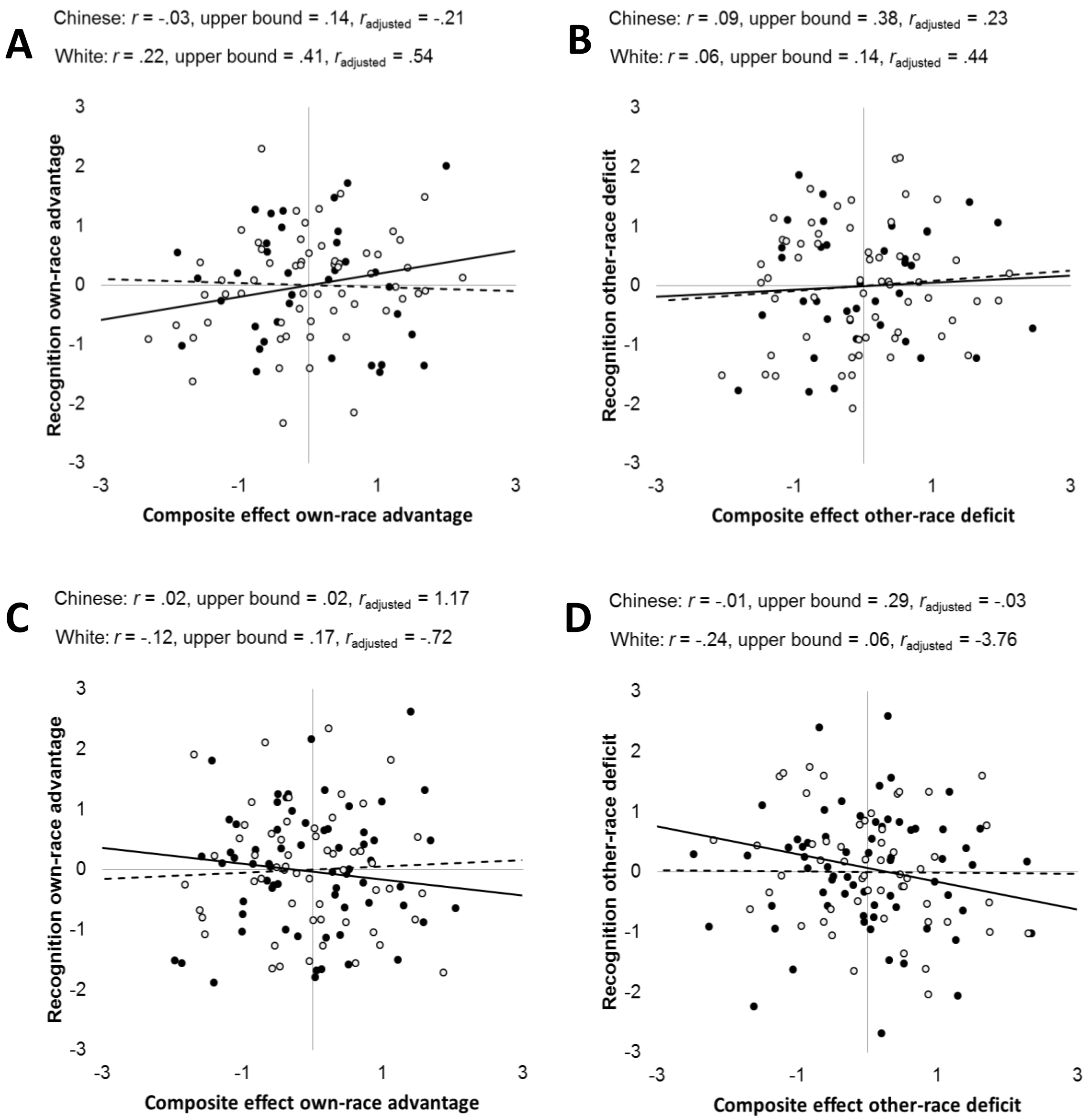

Figure 6. Correlations between the own-race effects in the composite task and in the recognition task in Experiments 1 (A and B) and 2 (C and D). Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.

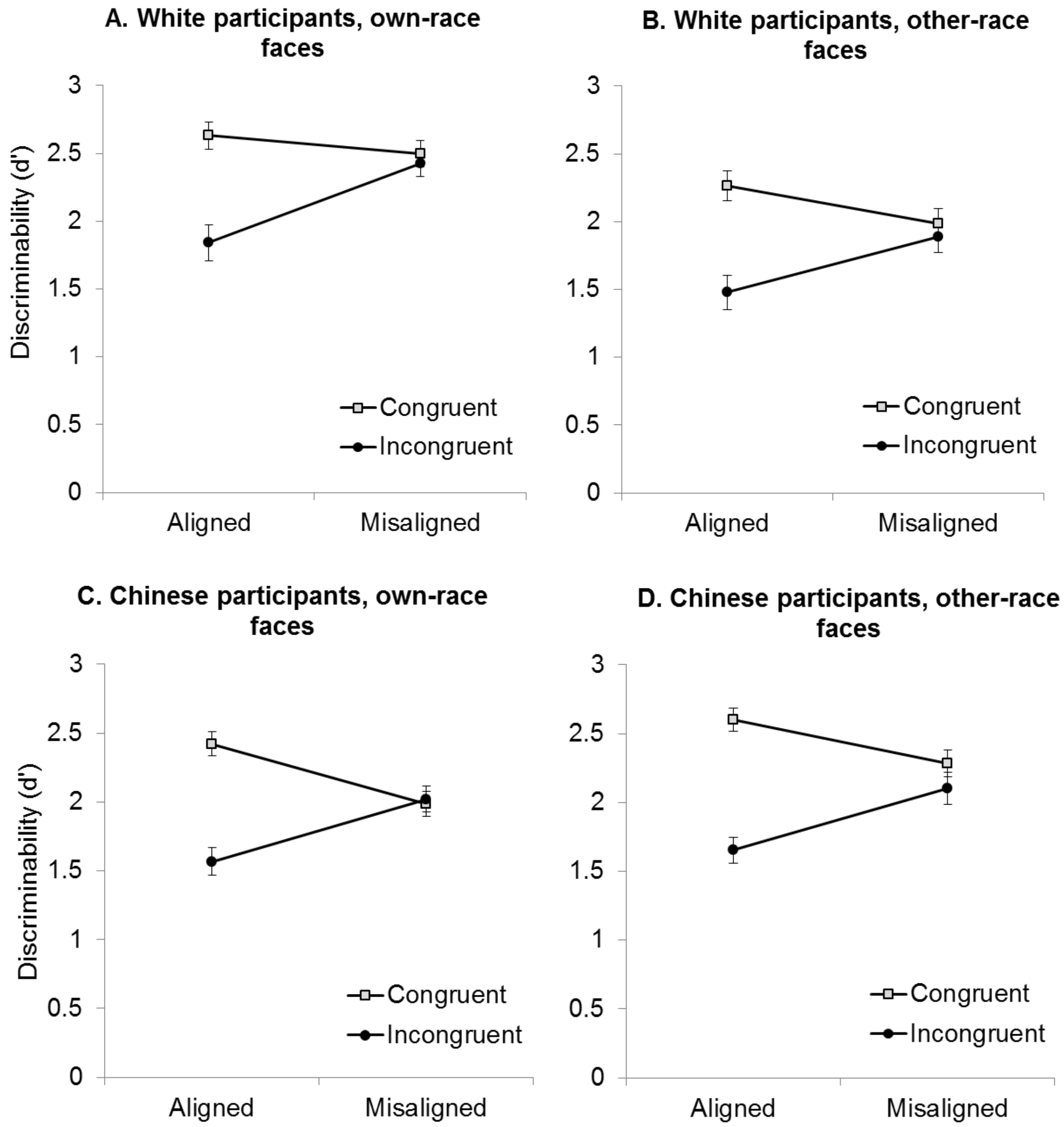

Figure 7. Discriminability (d') in the composite task in Experiment 3. Error bars represent standard error of the mean.

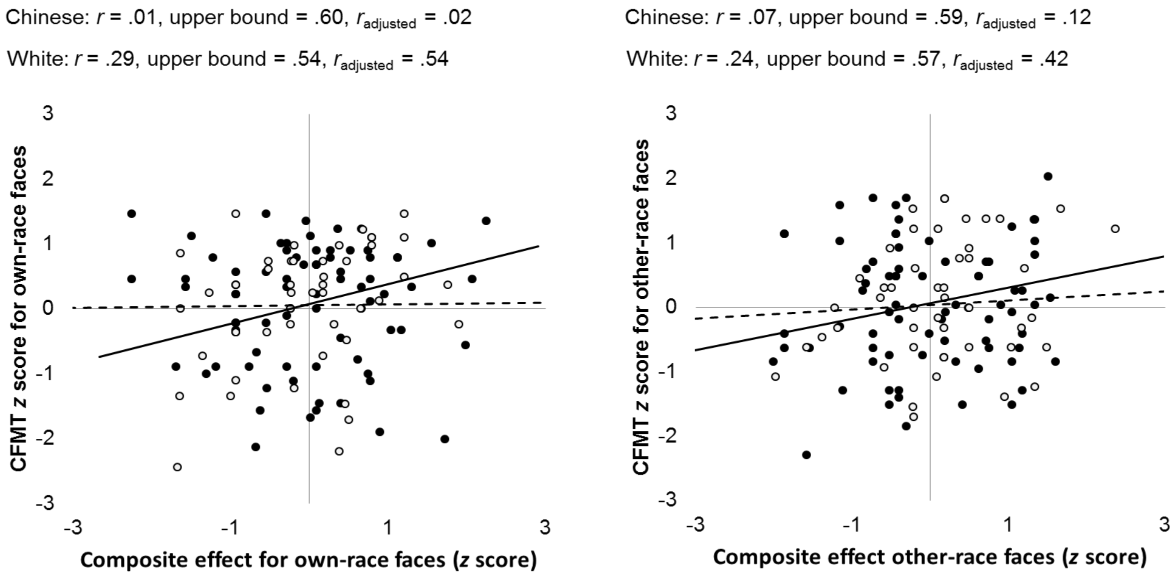

Figure 8. Within-race correlations between the composite effect and recognition in Experiment 3. Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.

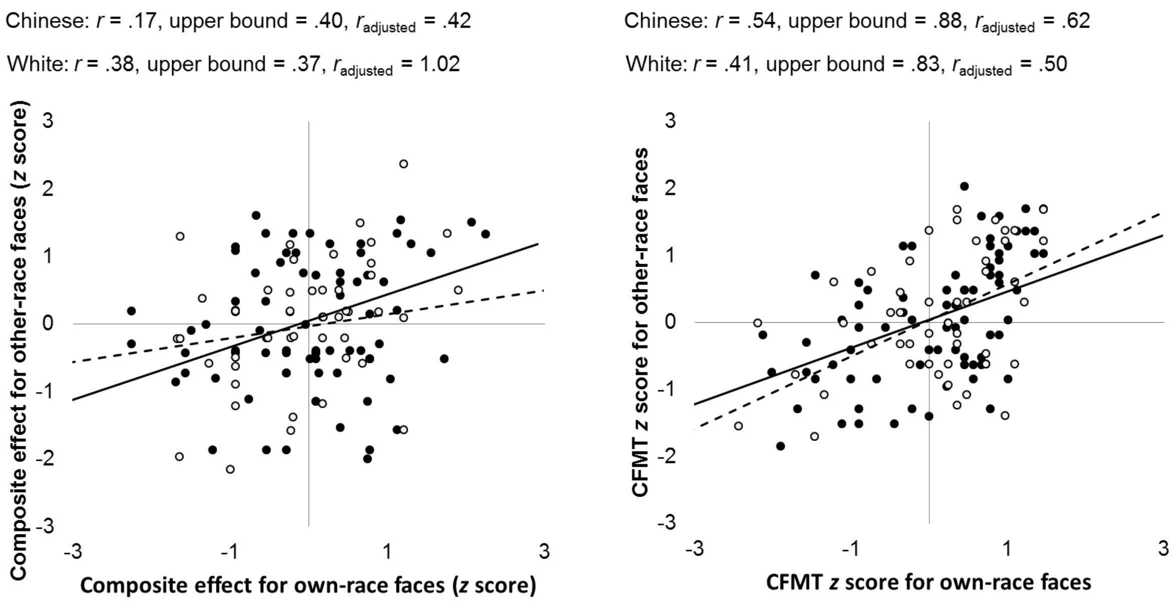

Figure 9. Within-task correlations between performance with own- and other-race faces in Experiment 3. Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.

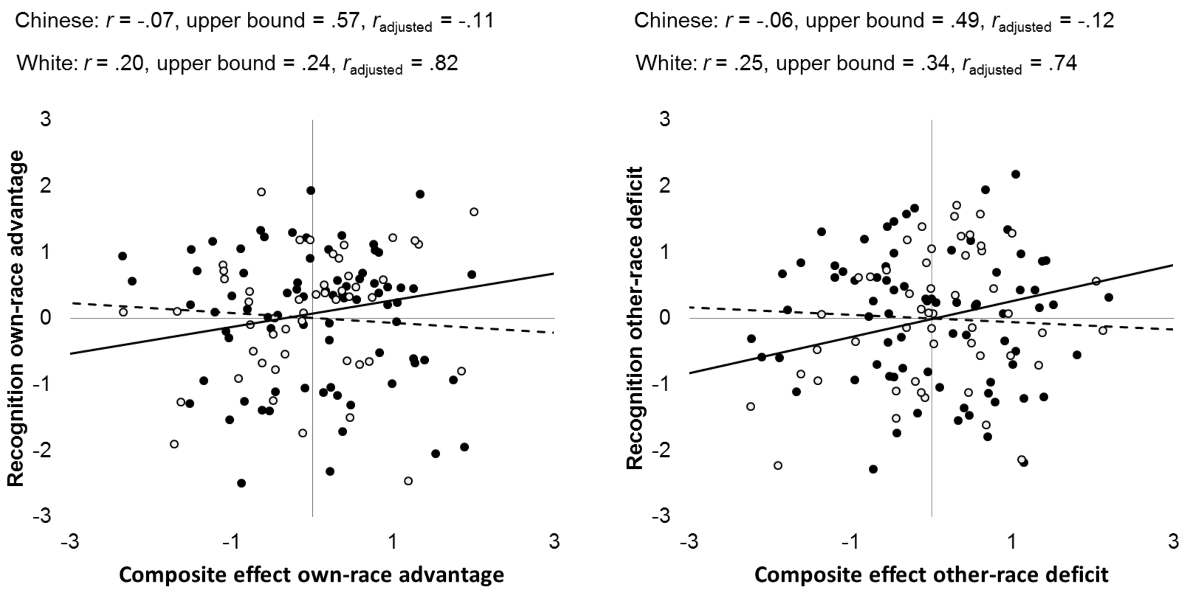

Figure 10. Correlations between the own-race effects in the composite task and recognition in Experiment 3. Dark circles represent Chinese participants; light circles represent White participants. The dashed line represents the line of best fit for Chinese participants; the solid line represents the line of best fit for White participants.