Perceptual Experience Shapes our Ability to Categorize Faces by National Origin: A New Other-Race Effect

Bianca Thorup[1,2], Kate Crookes[1,2], Paul P.W. Chang[3], Nichola Burton[1,2], Stephen Pond[1,2], Tze Kwan Li[4], Janet Hsiao[4], & Gillian Rhodes[1,2]

[1]ARC Centre of Excellence in Cognition and its Disorders [2]School of Psychological Science, University of Western Australia [3]School of Arts and Humanities, Edith Cowan University [4] School of Psychological Science, University of Hong Kong

Author Note This research was supported by the Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders (CE110001021) and an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300). We thank Sophie Joyce for creating the face morphs.

*Requests for reprints should be addressed to Gillian Rhodes, Department of Psychological Science, University of Western Australia, 35 Stirling Highway Crawley, WA 6009, Australia (email:gillian.rhodes@uwa.edu.au)

Abstract

People are better at recognising own-race than other-race faces. This other-race effect has been argued to be the result of perceptual expertise, whereby face-specific perceptual mechanisms are tuned through experience. We designed new tasks to determine whether other-race effects extend to categorising faces by national origin. We began by selecting sets of face stimuli for these tasks that are typical in appearance for each of six nations (three Caucasian, three Asian) according to people from those nations (Study 1). Caucasian and Asian participants then categorized these faces by national origin (Study 2). Own-race faces were categorized more accurately than other-race faces. In contrast, Asian-American participants, with more extensive other-race experience than the first Asian group, categorized other-race faces better than own-race faces, demonstrating a reversal of the other-race effect. Therefore, other-race effects extend to the ability to categorize faces by national origin, but only if participants have greater perceptual experience with own-race, than other-race faces. Study 3 ruled out non-perceptual accounts by showing that Caucasian and Asian faces were sorted more accurately by own-race than other-race participants, even in a sorting task without any explicit labelling required. Together, our results demonstrate a new other-race effect in sensitivity to national origin of faces that is linked to perceptual expertise.

Keywords: face perception, other-race effect, categorisation task, experience, perceptual expertise

Perceptual Experience Shapes our Ability to Categorize Faces by National Origin: A New Other-Race Effect

The other-race[1] effect (also known as the own-race effect, cross-race effect, own-race bias, and own-race advantage), is a well-established phenomenon, whereby people demonstrate better recognition of own-race, compared to other-race, faces (e.g., Malpass & Kravitz, 1969; for a meta-analysis see Meissner & Brigham, 2001). This phenomenon has been demonstrated in both laboratory and field settings, using a variety of tasks including matching and recognition memory tasks (e.g., Brigham, Maass, Snyder, & Spaulding, 1982; Lindsay, Jack, & Christian, 1991; Malpass & Kravitz, 1969; Meissner, Susa, & Ross, 2013; Slone, Brigham & Meissner, 2000; Wright, Boyd, & Tredoux, 2001).

The other-race effect in recognition has been shown across a diverse range of racial groups, which suggests that it may be universal (e.g., Ge et al., 2009; Malpass & Kravitz, 1969; O'Toole, Deffenbacher, Valentin, & Abdi, 1994; Wright et al., 2001). Additionally, researchers have demonstrated a crossover interaction between the race of the observer and race of face. For example, Caucasian participants demonstrate better recognition of Caucasian faces than Asian faces, and Asian participants demonstrate better recognition of Asian faces than Caucasian faces (O'Toole et al., 1994). This interaction suggests that the other-race effect is not due to differences in facial homogeneity between racial groups (Walker & Tanaka, 2003).

Many accounts of the other-race effect have been proposed (Sporer, 2001), but one long-held explanation is that we have greater perceptual expertise with own-race

[1] Following common practice, we use the term "race" to refer to visually distinct groups. These groups may nevertheless consist of individuals from a range of distinct ethnicities.

than other-race faces (Lebrecht, Pierce, Tarr, & Tanaka, 2009; Rhodes, Brake, Taylor, & Tan, 1989; Tanaka, Heptonstall, & Hagen, 2013; Tanaka, & Pierce, 2009). Support for this view comes from many studies that have found greater sensitivity to featural and configural differences between own-race than other-race faces (for a review, see Hayward, Crookes, & Rhodes, 2013). This perceptual expertise account argues that we develop expertise for own-race faces through experience, which tunes the underlying perceptual mechanisms required for face processing and recognition to the types of faces that we often experience (de Heering, de Liedekerke, Deboni, & Rossion, 2010; Feinman & Entwisle, 1976; Lee, Anzures, Quinn, Pascalis, & Slater, 2011; Rhodes, Hayward, & Winkler, 2006; Rossion & Michel, 2011; Telzer et al., 2013). In contrast, more limited experience with other racial groups reduces the development of expertise for other-race faces, resulting in poorer discrimination and recognition of other-race faces (de Heering et al., 2010; Feinman & Entwisle, 1976; Lee et al., 2011; Rhodes et al., 2006; Rossion & Michel, 2011; Telzer et al., 2013).

Given the importance of experience in shaping the face processing system and developing perceptual expertise for faces, one would expect to see a link between other-race social contact and the other-race effect. However, these studies have yielded mixed results. Some have shown no correlation (Brigham & Barkowitz, 1978; Carroo, 1987; Malpass & Kravitz, 1969; Ng & Lindsay, 1994) and Meissner and Brigham’s (2001) meta-analysis revealed only a small effect of social contact on the other-race effect. Hancock and Rhodes’ (2008) findings were consistent with this meta-analysis. However, others have found an association between higher other-race social contact and a reduced other-race effect (Chiroro & Valentine, 1995; Zhao, Hayward, & Bülthoff, 2014).

Importantly, however, extensive experience with other racial groups can eliminate or even reverse the other-race effect. For example, Asian adults who were born and raised in Australia or another Western country demonstrated no recognition advantage for Asian over Caucasian faces (Wan, Crookes, Reynolds, Irons, & McKone, 2015). Likewise, African American-born university students in the United States demonstrated no recognition advantage for African American over Caucasian faces (Lindsay et al., 1991). Additionally, Asian children and adolescents who were adopted by Western European families in early childhood demonstrated no other-race effect (de Heering et al., 2010). Finally, Korean adults, adopted by Western European families in childhood, demonstrated a reversal of the other-race effect (Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). These results suggest that perceptual expertise for other-race faces can be acquired through extensive experience with other racial groups. They also highlight the role of experience in the other-race effect, and support the perceptual expertise account.

In contrast to the large body of research examining the other-race effect in face recognition, there is little research on how face race affects face categorisation. Categorisation behaviour has played an important role in understanding the nature of perceptual expertise for objects more broadly, with experts able to make faster, more accurate categorisations at finer levels of discrimination than non-experts (Palmeri & Gauthier, 2004), and may therefore also be linked to face expertise. Thus far, researchers have focused on categorisation of faces by race (e.g., as Asian or Caucasian), finding that people categorize *other*-race faces by race faster than own-race faces, demonstrating a reversal of the other-race effect (i.e., the "other-race advantage"; Caldara, Rossion, Bovet, & Hauert, 2009; Ge et al., 2009; Levin, 1996, 2000; Valentine & Endo, 1992; Zhao & Bentin, 2008). However, this task does not tap expertise for own-race faces, and reduced perceptual expertise for other-race faces would lead to exactly this result, because it is associated with reduced sensitivity to differences between faces, which should make it easier to group faces, (i.e., to classify them by race).

Here we consider a different kind of categorisation that should be enhanced by sensitivity to differences between faces. Specifically, we asked participants from two races (Asian and Caucasian) to categorize own-race and other-race faces by national origin. The faces came from three (predominantly) Caucasian and three (predominantly) Asian nations. This task requires perceptual expertise, because the perceiver must use subtle visual cues to distinguish faces from the different nations. Thus, to the extent that the other-race effect reflects differences in face expertise, we would expect to see poorer categorisation of other-race than own-race faces.

For any accuracy to be possible, it is important that the faces from the different nations used are visually distinct and typical in appearance for their nations. Therefore, in Study 1, we obtained typicality ratings for faces from China, Japan, South Korea, Denmark, Germany and Italy, countries we deemed likely to contain such faces, rather than nations populated historically and/or recently by diverse immigration from many nations (e.g., Australia, USA). Participants from each nation rated the typicality of faces from their own nation. High agreement within each group would show that there are faces deemed to be typical for each nation, and thus potentially distinct in appearance from faces from other nations. This would then allow us to select the most typical faces for the national origins task used in Study 2.

In Study 2, we examined the performance of Caucasian and Asian participants, with extensive own-race, and more limited other-race, experience, on categorising own-race and other-race faces by national origin. Our main aim was to determine whether these groups would demonstrate an other-race effect, with better performance for own-race than other-race faces. According to the perceptual expertise account participants should demonstrate this effect, given their greater experience with own-race than other-race faces. To further explore the effect of experience, we recruited a second group of Asian participants who were either born and raised in the United States or had lived there for at least 8 years, and therefore had much more extensive other-race experience than the first Asian group. A reduced, or even reversed, other-race effect in this Asian-American group would strongly link performance to experience, consistent with the perceptual expertise account.

We presented faces both with and without hair visible. An other-race effect when hair is visible could simply reflect less knowledge about variations in hair colour, texture, and/or style for other-race people, and might tell us little about face expertise. Our primary interest was, therefore, in whether an other-race effect would be found when hair was not visible.

We were also curious about people’s confidence in their ability to categorize faces by their national origins. We therefore obtained confidence ratings before completing the national-origins categorisation task. It is an open question whether people think they can classify faces by their national origin and, if so, whether they are more confident for own-race faces, than other-race faces.

To foreshadow, we found an other-race effect for national origin categorisation in participants with little other-race experience in Study 2. In Study 3, we sought to rule out non-perceptual accounts of this finding, such as confusion about which faces belong to which nation, that is, a labelling problem. We asked Caucasian and Asian participants to sort the faces of each race (own-race, other-race) into three distinct, but unlabelled, national origin groups. This task has no labelling component and relies entirely on perceiving similarities between faces from the same nation and differences between faces from different nations. An other-race effect in this sorting task would, therefore, reflect differences in perceptual expertise, and reduce the plausibility of a non-perceptual account of the other-race effect observed in Study 2.

In Study 3, we also measured the size of the other-race effect in recognition memory, to explore whether there is any relationship between this classic other-race effect and any other-race effect on the national-origins sorting task.

Study 1

In Study 1, we asked participants from each of six nations to view faces of their own nation and rate their typicality, to determine whether there was within-group agreement, and identify the most typical faces of each nation for use in the national origins task in Study 2.

Method

Participants

Sixty adults (aged 21 to 40 years, *M* = 29.0 years, *SD* = 5.9 years), who identified themselves and both of their biological parents as Chinese, Japanese, South Korean, Danish, German or Italian were recruited online through social contacts. Ten were Chinese (5 male, 5 female; mean age = 22.5 years, *SD* = 2.8 years), ten were Japanese (4 male, 6 female; mean age = 30.9 years, *SD* = 5.28 years), ten were South Korean (4 male, 6 female; mean age = 28.5 years, *SD* = 4.48 years), ten were Danish (2 male, 8 female; mean age = 27.5 years, *SD* = 6.6 years), ten were German (4 male, 6 female; mean age = 31.5 years, *SD* = 6.36 years), and ten were Italian (3 male, 7 female; mean age = 32.8 years, *SD* = 3.33 years). All were born in, and resided in, the country they identified with until adulthood. The length of time participants reported living in a country outside of their country of birth ranged from 0 to 11 years (*M* = 18.8 months, *SD* = 27.3 months).

Stimuli and Materials

Five hundred and fifty-three (296 Asian, 257 Caucasian) front-view colour photographs of young adult athletes' faces, taken from the 2012 London Olympics website (http://www.bbc.com/sport/olympics/2012, November 2015)[2], were used for the typicality rating task. We selected relatively obscure athletes, intended to be unfamiliar to our participants. The Asian stimuli (146 male, 150 female) consisted of faces from China (50 male, 50 female), Japan (46 male, 50 female) and South Korea (50 male, 50 female). The Caucasian stimuli (122 male, 135 female) consisted of faces from Denmark (31 male, 35 female), Germany (50 male, 50 female) and Italy (41 male, 50 female). The athletes were photographed with a neutral expression, without glasses, facial hair or visible make-up, against a light background. All images were displayed at a resolution of 72 pixels inch. The face images displayed were 2 x 2.5 inches in width and height, respectively.

Procedure

The experiment was run online using Qualtrics Survey Software. Participants completed the rating task individually on their own computers (the task could not be completed on mobile or tablet devices). Participants were asked to provide demographic information, including their age, sex, nationality, country of birth and time spent living in a country outside of their birthplace. Participants born outside of the country they identified with could not proceed to the rating task.

In the rating task, participants viewed athletes' faces with the same national origin as themselves (e.g., Chinese participants looked at Chinese face stimuli). Participants saw two test blocks (one male and one female), with the order

[2] Photographs of Olympic athletes faces were chosen, as their national origin can be validated and the stimuli are of reasonable quality and are easily available.

counterbalanced across participants. In each test block, participants saw between 31 and 50 faces, shown individually, in an individually randomised order for each participant. Each face appeared on the screen with no time restriction. Participants were asked to rate the typicality of each face (e.g., rate how typically Chinese this face is) on a 9-point scale (“1 = Not at all”, “9 = Extremely”). Participants were also asked to indicate if they recognised the face, and those who answered yes were asked whether they could name the person or provide any information about them, so that any familiar faces could be excluded from the national origins task. The next face immediately appeared after their response. Participants were required to answer all questions. The experiment took on average 32 minutes to complete.

Results and Discussion

A background check was carried out on the athletes, and the faces of athletes who were born in a different country, or whose parents were born in a different country to the one they represented, were removed. In addition, the faces of athletes recognised by more than two participants were also removed. In total thirty-nine faces were removed from the stimuli set prior to the analyses (3 Chinese, 8 Japanese, 4 South Korean, 9 Danish, 8 German, 7 Italian), leaving 97 Chinese, 88 Japanese, 96 South Korean, 57 Danish, 92 German, and 84 Italian face stimuli remaining. The faces were then ranked in order of typicality for each sex, from each nation, based on participants’ responses. The mean and standard deviation of typicality ratings for all male and all female faces from each nation, and for the top 15 most typical faces from each nation are shown in Table 1.

The Chinese male, Danish female and male, German female and male, Italian female and male, Japanese female and male, and South Korean female typicality ratings all had high reliabilities, all Cronbach’s αs > .74 (see Table 1). The Chinese female and

South Korean male ratings were less reliable, Cronbach's α = .50 and α = .66 respectively, but still within the moderate range of reliability. Thus, participants agreed on typicality and we chose the 15 most typical faces for each nation, for each sex, to use for the national origins task in Study 2.

Table 1

Means (M) and Standard Deviations (SD) of the typicality ratings and Cronbach's α for all faces from each nation and sex. Means and Standard Deviations of the typicality ratings for the top 15 most typical faces from each nation and sex are also shown. N = number of faces.

Nation	Sex	*N*	Overall *M*	Overall *SD*	Cronbach's α	*N*	Top 15 *M*	Top 15 *SD*
Chinese	F	50	5.5	0.8	.50	15	6.5	0.4
Chinese	M	47	5.6	1.1	.74	15	6.7	0.4
Japanese	F	44	5.1	1.3	.85	15	6.4	0.5
Japanese	M	44	5.1	1.1	.80	15	6.3	0.6
South Korean	F	49	4.2	1.6	.90	15	6.2	0.7
South Korean	M	47	4.3	1.1	.66	15	5.4	0.5
Danish	F	30	5.9	1.2	.80	15	6.8	0.7
Danish	M	25	5.9	1.3	.81	15	6.6	1.1
German	F	46	5.9	1.1	.80	15	7.0	0.5
German	M	46	5.4	1.3	.84	15	6.8	0.8
Italian	F	45	5.2	1.4	.89	15	6.7	0.6
Italian	M	39	5.1	1.3	.86	15	6.3	0.3

Figure 1 shows an averaged composite face of each sex for each nation. These were made by morphing the 15 faces of each sex for each nation using standard morphing software (Psychomorph). They were not used as stimuli in Study 2, but are presented here to confirm that there were consistent visual differences between faces from the three nations of each race, a necessary condition for any accuracy on the national origins categorisation task in Study 2.

Figure 1. Face morphs created from the top 15 most typical faces of Denmark, Germany, Italy, China, Japan and South Korea, as rated by people from those nations in the typicality ratings task.

Study 2

In Study 2, we examined the performance of Caucasian, Asian and Asian-American participants on categorising own-race and other-race faces by national origin to determine whether the other-race effect extends to this categorisation task, and to examine the effects of experience on any observed other-race effect.

Method

Participants

One hundred and twenty-four Caucasian participants were recruited online through Amazon Mechanical Turk (MTurk) and received US$1 in return for their participation. All Caucasian[3] participants were born in and resided in the United States (US), and had spent less than two years living in an Asian country. Fifty-nine Asian

[3] All Caucasian participants identified as white Caucasian, and all Asian participants identified as East or South-East Asian.

participants were recruited. Of these, seven were recruited online through social contacts (4 lived in Asia, 3 lived in Australia) and received no payment for their participation, and 52 were recruited online from (information removed for blind review) and either received AU$10 for their participation or a course credit. All the university recruits had spent less than two years living in Australia or another predominantly Caucasian country. In addition, a second group of 53 Asian participants with more extensive other-race (Caucasian) experience was recruited online through MTurk, and received US$1 in return for their participation. These Asian-American participants were either born in the US or had lived there for more than 8 years, and all currently resided in the US.

The data from participants who reported visual or computer problems, or did not enter valid responses (i.e., repeatedly clicking the same response) were removed. In addition, the data from participants aged over 50 years were also removed. This upper age limit was selected based on the results of a study by Germine, Duchaine, and Nakayama (2011), which showed poorer performance for participants aged over 50 years than for young adults on the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). The final group of 210 participants consisted of 104 Caucasian (49 male, 55 female, aged 20 to 50 years, M = 32.2 years, SD = 7.5 years), 54 Asian (21 male, 33 female, aged 17 to 31 years, M = 22.1 years, SD = 3.6 years), and 52 Asian-American (25 male, 27 female, aged 18 to 49 years, M = 30.2 years, SD = 7.7 years) participants. None of these participants lived in any of the countries in the national origins test (except for four Asian Participants, three who lived in China and one who lived in Korea).

The differences in sample sizes between the groups are due to the different sampling methods. Importantly, a power analysis indicated that all groups met the minimum sample size of 34 per group required to detect a medium-sized other-race effect (d = 0.5) with 80% power with alpha at .05. We based our expectation of a medium sized effect on the effect size for the other-race effect in face recognition (Meissner & Brigham, 2001).

Stimuli and Materials

One hundred and eighty (90 Asian, 90 Caucasian) front view colour images of young adult athletes' faces were used for the national origins task. The 90 Asian face stimuli consisted of 30 (15 male, 15 female) typical faces each from China, Japan and South Korea. The 90 Caucasian face stimuli consisted of 30 (15 male, 15 female) typical faces each from Denmark, Germany and Italy. The images selected were deemed the most typical faces of their nation, according to the results of Study 1.

Faces were edited using Adobe Photoshop CS3 ™. All faces were standardised to have horizontally aligned pupils and an interpupil distance of 50 pixels. For the "with-hair" condition the face images were cropped using the magnetic lasso tool, leaving the hair, face, ears and top of the neck visible, and removing clothing cues. These faces were then masked using the elliptical marquee tool so that most of the hair was excluded for the "hair-masked" condition. All images were displayed at a resolution of 72 pixels inch on a grey background. The face images displayed were 2 x 2.5 inches in width and height, respectively.

Procedure

The experiment was run online using Qualtrics Survey Software. Participants completed the national origins task on their own computers (the task could not be completed on mobile or tablet devices). They were asked to provide demographic information, including their age, sex, nationality, country of birth, and time spent living in a country outside of their birthplace. Participants were also asked how confident they were that they could categorize an Asian face as Chinese, Japanese or South Korean, and a Caucasian face as Danish, German or Italian on a 9-point scale (1 = ‘Not at all’, 9 = ‘Extremely’), before completing the national origins task.

In the national origins task we examined participants’ performance for categorising own-race and other-race faces by national origin using a three-alternative forced choice (3AFC) procedure. Participants saw two test blocks of faces, one with hair visible and one with hair masked, with the order counterbalanced across participants. In each test block, participants saw 180 faces (90 of each race, 45 male, 45 female), presented individually, in an individually randomised order for each participant. Each face remained on the screen until the national origin of the face was selected from the three options presented on the screen (e.g., Chinese, Japanese, South Korean, or Danish, German, Italian) using the mouse. The next face appeared immediately after the selection was made. The experiment took an average of 43 minutes to complete.

Results and Discussion

Accuracy

Categorisation accuracy (percent correct) was calculated for own-race and other-race faces in the with-hair and hair-masked conditions (Figure 2). A three-way repeated measures analysis of variance (RM ANOVA) was conducted on accuracy (percent correct), with participant group (Caucasian, Asian, Asian-American) as a between-participants variable, and face race (own-race, other-race) and hair (with-hair, hair-masked) as within-participant variables. Planned t-tests were used to follow up interactions, where appropriate.

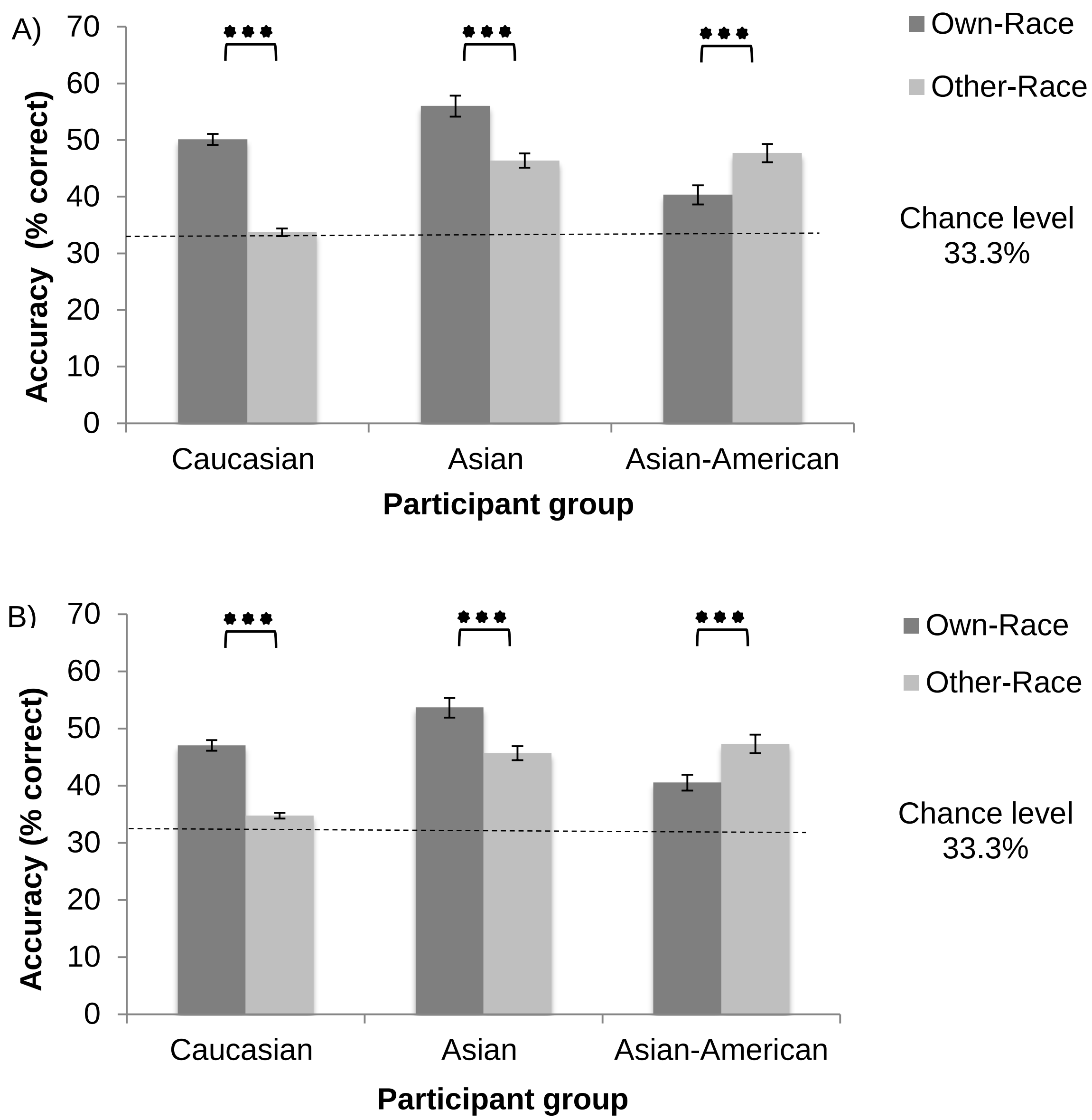

Figure 2. Mean accuracy for categorising own-race and other-race faces by national origin, shown (A) with-hair, and (B) with hair-masked, for each participant group. Error bars show ± 1 SEM. $p < .001$***, $p < .01$**

An initial four-way RM ANOVA with block order (hair first, hair-masked first) as an additional between-participants variable revealed no significant main effect of block order or any significant interactions involving block order, all Fs < 2.50, all ps > .09, all η_p^2s < .02. Therefore, block order was removed from further analyses.

The 3-way ANOVA yielded significant main effects for face race, $F(1, 207) = 41.62$, $p < .001$, $\eta_p^2 = .17$, hair, $F(1, 207) = 5.55$, $p = .019$, $\eta_p^2 = .03$, and participant group, $F(2, 207) = 28.20$, $p < .001$, $\eta_p^2 = .21$. However, these effects were qualified by interactions between participant group and face race, $F(2, 207) = 60.98$, $p < .001$, $\eta_p^2 = .37$, and face race and hair, $F(1, 207) = 5.74$, $p = .017$, $\eta_p^2 = .03$. There was no interaction between hair and participant group, $F(2, 207) = 1.05$, $p = .35$, $\eta_p^2 = .01$). There was also a significant 3-way interaction between participant group, face race and hair, $F(2, 207) = 4.03$, $p = .019$, $\eta_p^2 = .04$ (Figure 2). To explore this three-way interaction, we conducted separate two-way RM ANOVAs with face race and hair as within-participant variables for each participant group.

Caucasian participants

There were significant main effects of face race, $F(1, 103) = 231.35$, $p < .001$, $\eta_p^2 = .69$, and hair, $F(1, 103) = 4.72$, $p = .03$, $\eta_p^2 = .04$. There was also a significant interaction between face race and hair, $F(1, 103) = 17.18$, $p < .01$, $\eta_p^2 = .14$. Importantly, however, planned paired-samples *t* tests showed that performance was better for own-race than other-race faces when hair was masked, $t(103) = 12.68$, $p < .001$, $d = 1.24$ (own-race $M = 47.0\%$, $SD = 9.3\%$; other-race $M = 34.7\%$, $SD = 5.2\%$), as well as when hair was visible (own-race $M = 50.1\%$, $SD = 9.9\%$; other-race $M = 33.7\%$, $SD = 6.9\%$), $t(103) = 14.25$, $p < .001$, $d = 1.40$. Therefore, the other-race effect occurred with face cues alone, as expected if it reflects differences in face expertise.

Accuracy was significantly greater for own-race faces shown with hair ($M = 50.1\%$, $SD = 9.9\%$), than with hair masked ($M = 47.0\%$, $SD = 9.3\%$), $t(103) = 4.28$, $p < .001$, $d = .42$. In contrast, there was no significant difference in accuracy for other-race faces shown with hair ($M = 33.7\%$, $SD = 6.9\%$) and with hair masked ($M = 34.7\%$, $SD =$

5.2%), *t*(103) = -1.60, *p* = .113, *d* = -.16. Therefore, participants were able to use hair cues for own-race, but not other-race, faces.

Asian participants

There was a significant main effect of face race, *F*(1, 53) = 27.01, *p* < .001, η_p^2 = .34, with greater accuracy for own-race faces (*M* = 54.8%, *SD* = 12.6%) than other-race faces (*M* = 46.0%, *SD* = 8.2%). There was no significant main effect of hair, *F*(1,53) = 3.64, *p* = .062, η_p^2 = .064, and no interaction, *F*(1, 53) = 1.61, *p* = .210, $\eta_p^2 = .03$.

Asian-American participants

There was a significant main effect of face race, *F*(1, 51) = 14.08, *p* < .001, η_p^2 = .22, with significantly *lower* accuracy for own-race (*M* = 40.4%, *SD* = 10.7%), than other-race faces (*M* = 47.5%, *SD* = 11.0%). There was no significant main effect of hair, *F*(1, 51) = .02, *p* = .890, η_p^2 < .01, and no interaction, *F*(1,51) = .22, *p* = .645, η_p^2 = .01.

In summary, both Caucasian and Asian participants categorized own-race faces by national origin significantly better than other-race faces, demonstrating an other-race effect. However, as expected, Asian-American participants did not show the same effect. Rather, they categorized other-race faces significantly better than own-race faces, demonstrating a reversal of the other-race effect. This result provides strong support for a link between experience and performance, as expected if the other-race effect results from differences in perceptual expertise. Importantly, the race effects were not dependent on the presence of hair cues, although Caucasian participants were able to use hair cues to improve performance for own-race faces. Therefore, the results cannot be attributed simply to greater familiarity with own-race than other-race hairstyles.

Accuracy Compared to Chance

One-sample *t* tests were used to determine whether accuracy (percent correct) was greater than chance (33.3%) for each participant group, face race and hair condition (Table 2). In all cases, performance was significantly above chance, except for Caucasian participants categorising other-race faces with hair visible.

Table 2

Mean accuracy for categorising own-race and other-race faces shown with hair and hair-masked compared to chance.

Race & Hair	*t*(df)	*p*	*d*	95% CI
Caucasian				
Own-race hair	17.23(103)	< .001	1.69	[14.9, 18.7]
Own-race hair masked	15.11(103)	< .001	1.48	[11.9, 15.5]
Other-race hair	0.62(103)	= .536	.06	[-.92, 1.75]
Other-race hair masked	2.85(103)	= .005	.28	[0.4, 2.4]
Asian				
Own-race hair	12.29(53)	< .001	1.67	[19.0, 26.4]
Own-race hair masked	11.76(53)	< .001	1.6	[16.9, 23.8]
Other-race hair	10.37(53)	< .001	1.41	[10.5, 15.6]
Other-race hair masked	10.20(53)	< .001	1.39	[10.0, 14.8]
Asian-American				
Own-race hair	4.13(51)	< .001	.57	[3.6, 10.4]
Own-race hair masked	5.21(51)	< .001	.72	[4.4, 10.0]
Other-race hair	8.94(51)	< .001	1.24	[11.2, 17.6]
Other-race hair masked	8.65(51)	< .001	1.20	[10.7, 17.2]

Confidence

A two-way RM ANOVA was conducted on confidence ratings, with participant group (Caucasian, Asian, Asian-American) as a between-participants variable, and face race (own-race, other-race) as a within-participants variable. The means and standard errors for all conditions are shown in Figure 3.

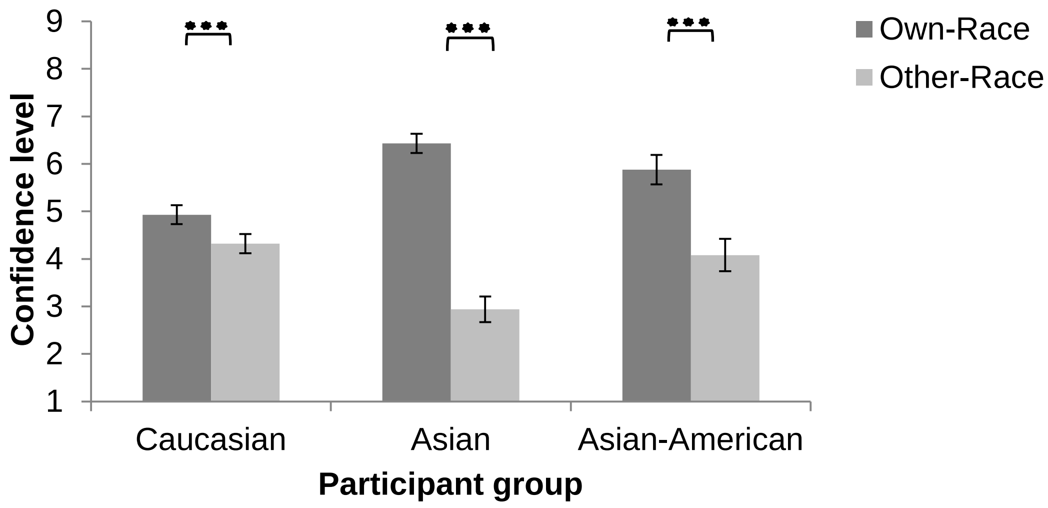

Figure 3. Mean self-reported confidence for categorising own-race and other-race faces by national origin for each participant group. Error bars show $\pm$ 1 SEM. $p < .001$ ***

There was a significant main effect of face race, $F(1, 207) = 185.77$, $p < .001$, $\eta_p^2 = .47$. There was no significant main effect of participant group, $F(2, 207) = 0.70$, $p = .500$, $\eta_p^2 <.01$. The interaction effect between face race and participant group was also significant, $F(1, 207) = 37.07$, $p < .001$, $\eta_p^2 = .26$. Inspection of Figure 3 suggests that confidence was lower for other-race faces than own-race faces for all groups, but that this difference was larger in the Asian and Asian-American groups than the Caucasian group.

Study 3

We asked Caucasian and Asian participants to sort Caucasian and Asian faces into three distinct (but unnamed) nations for each race, to determine whether an other-race effect is still obtained when the labelling component of the task is removed. We also examined performance on own-race and other-race face recognition, to check that our participants show the usual other-race effect in face recognition, and to determine whether there is any correlation between the other-race effects shown in these two different tasks.

Method

Participants

Sixty Caucasian participants were recruited from the (information removed for blind review) and either received AU$15 for their participation or course credit. All Caucasian participants were born and raised in Australia, and had spent less than two years living in a non-Western country. Sixty-six Asian participants were recruited from the (information removed for blind review) and received course credit for their participation. All Asian participants were born and raised in China (Hong Kong or mainland), and had spent less than two years living in a non-Asian country.

Participants who did not complete the sorting task correctly (i.e., placed an unequal number of face stimuli in the 3 piles) (N = 1, Asian; N = 6, Caucasian), or did not complete the CFMTs (N = 1, Caucasian) were excluded. The final group of 118 participants consisted of 53 Caucasian participants (20 male, 33 female), aged between 17 and 37 years (*M* = 22.8 years, *SD* = 5.3 years), and 65 Asian participants (32 male, 33 female), aged between 18 to 26 years, (*M* = 20.7 years, *SD* = 1.9 years). None of the Caucasian participants lived in any of the countries in the sorting tasks, in contrast to the Asian participants, who were all from Hong Kong (China).

Stimuli and Materials

The same 180 (90 Asian, 90 Caucasian) face images, used in Study 2, were used for the sorting task. The face images were printed on cards, 2 x 2.6 inches in width and height, respectively, on a grey background, and laminated.

Procedure

Participants completed the national-origins sorting task on a desk in a testing room. Participants sorted four sets of faces (Caucasian male, Caucasian female, Asian male, Asian female), with the order counterbalanced across participants. In each set, participants were presented with 45 face cards (15 from each of 3 nations), which were shuffled and arranged unsystematically on a large table. The following instructions were given verbally: “These faces belong to three nations. Please group them together by national origin. Please ensure that you make three groups of 15 faces. You have 5 minutes to perform this task. I will start the timer when I say begin. I will advise you when there is 1 minute remaining. The timer will sound when the time has elapsed. Do you have any questions? Please begin”.

After completing the national-origins sorting task, they were asked to complete a questionnaire using Qualtrics Survey Software, which asked them what six nationalities they thought the faces were. Participants were also asked how confident they were that they could sort Asian and Caucasian faces by national origin on a 9-point scale (1 = ‘Not at all’, 9 = ‘Extremely’).

After completing the sorting task, participants completed the Australian and Chinese CFMTs (McKone et al., 2011), with order counterbalanced across participants. Participants were seated approximately 50 cm in front of a desktop computer. The CFMT requires participants to learn 6 male target identities and then recognise them under increasingly difficult conditions. In phase one, for each target participants are shown the face in three different views, and then complete three test trials where they are required to identify the target (shown in the same viewpoints as learning) amongst two distractors. In phase two the target faces, again shown with two distractors on each trial, are presented under different lighting conditions and from different viewpoints. Lastly in phase three targets and distractors are presented with visual noise added to the images.

Finally, participants completed another questionnaire using Qualtrics Survey Software, to provide demographic information, including their age, sex, race, ethnic origin, country of birth, country of their parents' birth, and time spent living in a country outside of their birthplace.

Results and Discussion

National-Origins Sorting Task

We coded the national origin of all faces, and then considered all possible pairs of faces from the stimulus set. We measured performance using coherence values, calculated as the proportion of pairs from the same national origin that were correctly placed in the same category, minus the proportion of pairs from different national origins that were incorrectly placed in the same category. If all faces of the same national origin were sorted correctly into the same group, and all faces of different national origins were sorted correctly into different groups, then the coherence value would equal 1. If the faces were sorted randomly into groups, by chance, the coherence value would equal 0. A two-way repeated measures analysis of variance (RM ANOVA) was conducted on the coherence values, with participant group (Caucasian, Asian) as a between-participants variable, and race of face (own-race and other-race) as a within-participant variable. Planned t-tests were used to follow up interactions, where appropriate.

The 2-way ANOVA yielded significant main effects for face race, $F(1, 116) = 91.99$, $p < .001$, $\eta_p^2 = .44$ and participant group, $F(1, 116) = 13.91$, $p < .001$, $\eta_p^2 = .11$, and a significant interaction between the two, $F(1, 116) = 78.94$, $p < .001$, $\eta_p^2 = .41$ (Figure 4).

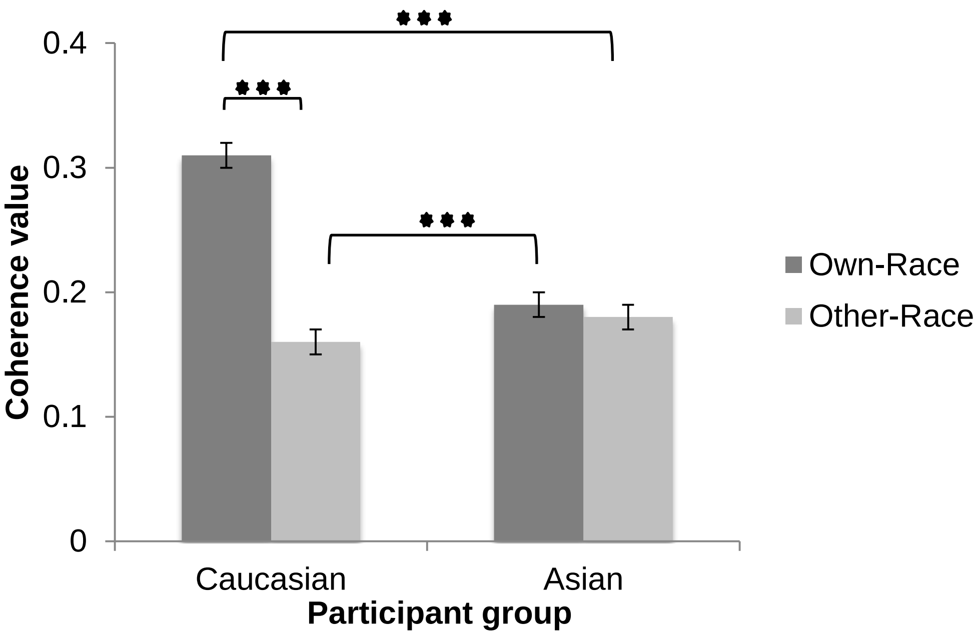

Figure 4. Mean coherence values for categorising own-race and other-race faces by national origin, for each participant group. Error bars show ± 1 SEM. $p < .001$***

Caucasian participants showed significantly higher coherence values for own-race faces, than other-race faces, $t(52) = 13.06$, $p < .001$, $d = 1.81$. In contrast, Asian participants showed no significant difference in coherence values for own-race faces and other-race faces, $t(64) = .51$, $p = .61$, $d = -0.06$. These results could potentially reflect a stimulus set effect. For example, greater variation (e.g., in hair and skin colouring) in the Caucasian face set may have offset any other-race effect for the Asian participants (and amplified it for Caucasian participants).

To avoid any such stimulus-set effects, we examined whether there was an other-race effect for each race of faces separately. Importantly, Caucasian faces were sorted significantly better by Caucasian than Asian participants, $t(116) = 7.67$, $p < .001$, $d = 0.71$, and Asian faces were sorted significantly better by Asian than Caucasian participants, $t(116) = 2.09$, $p = .038$, $d = 0.19$. These results indicate a genuine other-race effect for both groups on this perceptual sorting task.

Confidence

The same ANOVA was conducted on confidence ratings. Not surprisingly, participants were more confident about sorting own-race ($M = 5.6$, $SD = 1.6$) than other-race ($M = 3.8$, $SD = 1.5$) faces , $F(1, 116) = 110.16$, $p < .001$, $\eta_p^2 = .49$ (Figure 5). There was no significant main effect of participant group, $F(1, 116) = 0.12$, $p = .73$, $\eta_p^2 < .01$, and no interaction, $F(1, 116) = .04$, $p = .84$, $\eta_p^2 = .0$.

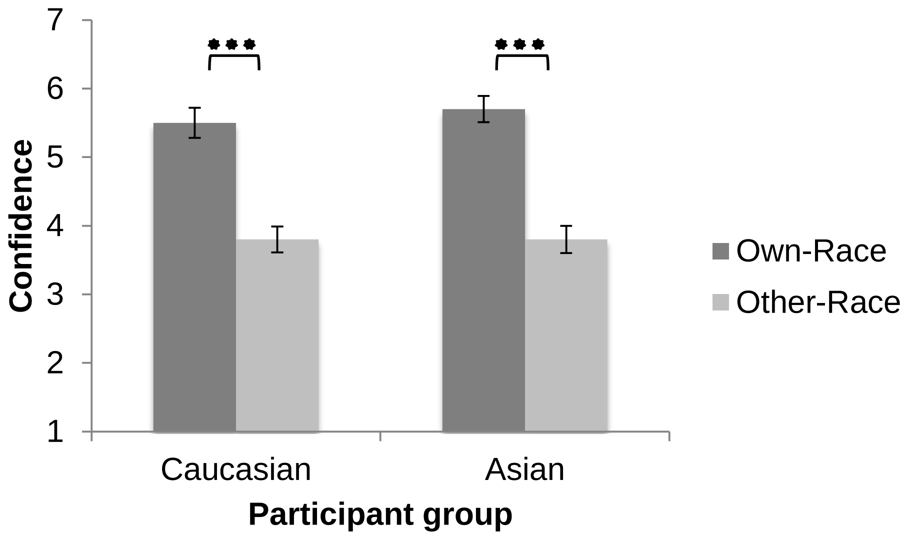

Figure 5. Mean self-reported confidence for sorting own-race and other-race faces by national origin for each participant group. Error bars show ± 1 SEM. $p < .001$ ***

Face recognition (CFMT) Performance

The same ANOVA was conducted on accuracy (proportion correct) on the CFMTs (Figure 6). As expected, performance was better for own-race ($M = .78$, $SD = .11$) than other-race ($M = .72$, $SD = .11$) faces, $F(1, 116) = 54.68$, $p < .001$, $\eta_p^2 = .32$, replicating the well-known other-race effect in face recognition memory. Caucasian participants (M = .78, SD = .11) performed better than Asian participants (M = .71, SD = .10), $F(1, 116) = 13.47$, $p < .001$, $\eta_p^2 = .10$, but there was no interaction between participant group and face race, $F(1, 116) = 3.03$, $p = 0.84$, $\eta_p^2 = .03$.

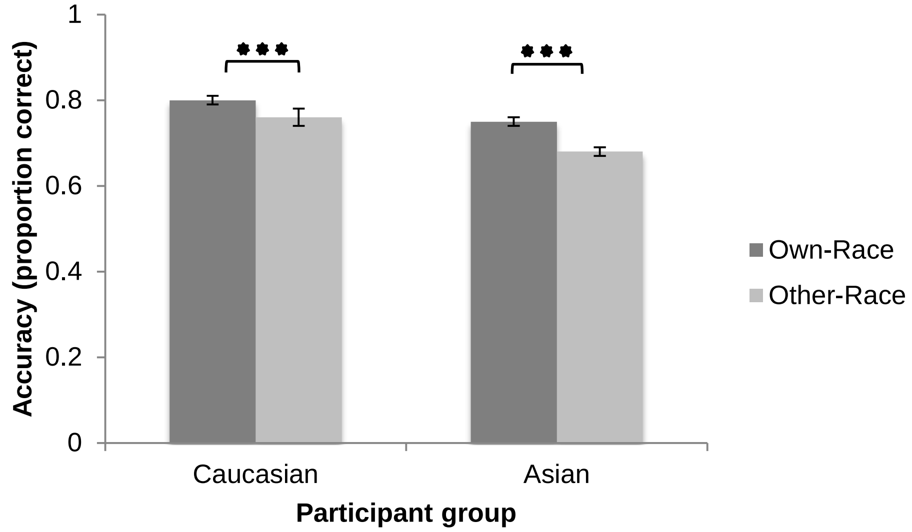

Figure 6. Mean accuracy for recognising own-race and other-race faces, on the Australian and Chinese Cambridge Face Memory Tests, for each participant group. Error bars show ± 1 SEM. $p < .001$***

Correlation between the other-race effect on the national-origins sorting task and the CFMTs

For the sorting task, the other-race effect was calculated by subtracting coherence values for categorising other-race faces from their coherence values for categorising own-race faces. For the CFMT, the other-race effect was calculated by subtracting accuracy (proportion correct) for recognition. The size of the other-race effect in the sorting task was not significantly related to the size of the other-race effect on the CFMT for either Caucasian participants, $r(51) = -0.08$, $p = .294$ “1-tailed”, or Asian participants, $r(63) = .15$, $p = .123$ “1-tailed”.

Identifying the three national origins for each race

The same ANOVA as above was conducted on participants' accuracy (proportion correct) for explicitly identifying the three national origins for each race.

There was no main effect for face race, $F(1, 116) = .04$, $p = 0.85$, $\eta_p^2 = 0.0$ (Figure 7). There was a significant main effect for participant group, $F(1,116) = 4.88$, $p = .03$, $\eta_p^2 = .04$. However, this effect was qualified by an interaction between face race and participant group, $F(1, 116) = 128.17$, $p < .001$, $\eta_p^2 = .53$.

Both Asian, $t(64) = 7.64$, $p < .001$, $d = 0.96$, and Caucasian, $t(52) = 9.12$, $p < .001$, $d = 1.26$, participants were significantly better at explicitly identifying the national origins of Asian faces (Chinese, Japanese, Korean), than Caucasian faces (Danish, German, Italian).

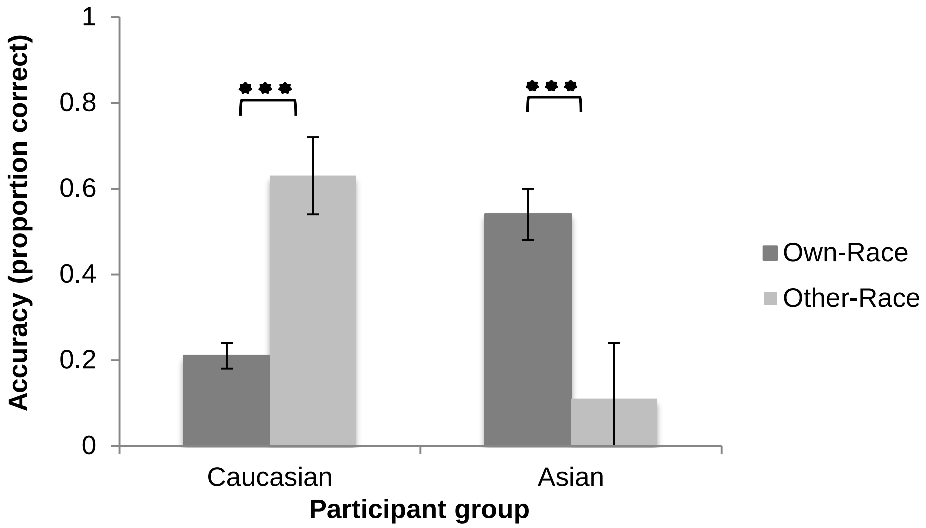

Figure 7. Mean accuracy for identifying the national origins of own-race and other-race faces for each participant group. Error bars show ± 1 SEM. $p < .001$***

General Discussion

In Study 1, the most typical faces of China, Japan, South Korea, Denmark, Germany and Italy were identified through the typicality ratings task. This task validated a set of face stimuli that are typical in appearance of these six nations that could be used to examine participants' performance for categorising own-race and other-race faces on the national origins task. In Study 2, we found evidence of an other-race effect in the national-origins categorisation task. Caucasian and Asian participants categorized own-race faces by national origin significantly better than other-race faces. In Study 3, we found evidence of an other-race effect in the national-origins sorting task. Caucasian and Asian faces were sorted more accurately by own-race, than other-race participants. Our results are the first to demonstrate that other-race effects extend to tasks that require participants to categorize faces by national origin.

An important finding in Study 2 was that in contrast to Asian participants, Asian-American participants, who were either born in the US or had spent more than 8 years in the US and had extensive other-race experience, demonstrated a reversal of the other-race effect. This result supports a role for perceptual expertise. Sangrigoli et al. (2005) reported a similar finding in the recognition literature with Korean adults who were adopted by Western European families in childhood. These researchers suggested that social interaction continues to shape the face processing system, and that being immersed in a predominantly Caucasian environment with limited own-race experience can reverse the other-race effect. Our results provide further evidence linking other-race effects with experience, and are consistent with a perceptual expertise account.

It is important to note that we found evidence to suggest that people can categorize own-race and other-race faces by national origin when hair is masked. This finding demonstrates that performance cannot be attributed to knowledge of hair appearance for different nations, but rather reflects differences in ability to discriminate subtle differences between faces. This finding also raises the question, how are people making these national origin judgements? All participants were able to categorize faces with hair masked above chance level, which suggests that they are using internal face cues to make these judgements. This result is not surprising, given that internal facial features are used to learn new faces and communicate social information (Longmore, Liu, & Young, 2015).

Hair did not appear to be an important cue to national origin. Only Caucasian participants showed any advantage for categorising faces with hair over faces with hair masked, and only for own-race faces. Neither Asian nor Asian-American participants showed any such advantage for categorising faces with hair.

The national-origins categorisation task required participants to label faces by nation. This requirement raises the possibility that our task does not tap perceptual expertise and that poorer performance with other-race reflects incorrect labelling. Perhaps participants are equally good at detecting subtle differences between faces of different nationalities, but incorrectly allocate the groups to countries. To address this possibility, we examined the proportion of faces of the same nation that were given the same label, regardless of the correctness of that label. This coherence analysis (presented in Supplemental Materials) revealed the same pattern of results as the accuracy analysis presented above, with a significant other-race effect for Caucasian and Asian participants (own-race faces better grouped by national origin than other-race faces) and a reversal of the other-race effect for Asian-American participants. Therefore, although our national-origins categorisation task requires participants to label the faces, we believe that it taps differences in perceptual expertise with own-race and other-race faces.

To more directly rule out non-perceptual, knowledge-based accounts, in Study 3 we removed the labelling component of the national-origins categorisation task and still found evidence of an other-race effect. Caucasian participants sorted own-race faces by national origin significantly better than other-race faces, demonstrating an other-race effect. Asian participants did not show the same effect, demonstrating similar performance for sorting own-race and other-race faces. This result suggests that Caucasian faces may be easier to sort (e.g., due to hair colour cues), which may ameliorate the other-race effect for Asian participants. Importantly, however, both Asian and Caucasian faces were sorted by national origin better by people of the same race than by people of another race. Therefore, we suggest that the other-race effect in Study 2 was most likely attributed to perceptual expertise, and not simply the result of incorrect labelling.

Not surprisingly, all groups were more confident in their ability to categorize own-race than other-race faces. These results mirror findings that people are more confident that they can discriminate own-race faces than other-race faces (Corenblum & Meissner, 2006; Meissner et al., 2013). It is important to note, however, that although Asian-American participants' confidence in their ability to categorize other-race faces was lower than for own-race faces, their accuracy for categorising other-race faces was higher than for own-race faces. This result suggests that they are not aware of their other-race face expertise, and indicates that confidence need not match accuracy, as is well known from the eyewitness literature (Sporer, Penrod, Read, & Cutler, 1995).

Caucasian and Asian participants also demonstrated the classic other-race effect in face recognition, which has been consistently shown in the literature (Meissner & Brigham, 2001). We did not find a correlation between the size of the other-race effect found in the CFMTs and national-origins sorting task. This dissociation might reflect the very different demands of the two tasks, with the CFMTs focusing on learning and memory, in contrast to the sorting task, which relies on perceptual discrimination.

All participants were more accurate at explicitly identifying the three national origins used for Asian faces in Study 3. Note, however that the Asian and Caucasian faces selected for this task were all from East Asia and Europe, respectively, and there are only eight countries in East Asia compared to 27 countries in Europe. This may have made it more difficult for people to identify the national origin of Caucasian faces. Moreover, we did not tell participants the geographical origin of faces, further increasing the possible number of Caucasian countries, and potentially reducing accuracy of identifying the three target nations.

Together, the results demonstrate a new other-race effect that cannot be attributed to non-perceptual factors, and provide support for the perceptual expertise account of other-race effects (Lebrecht et al., 2009; Rhodes et al., 1989; Tanaka et al., 2013; Tanaka, & Pierce, 2009), which argue that face expertise is shaped through experience (Lee et al., 2011; Rhodes, et al., 2006; Rossion & Michel, 2011; Telzer et al., 2013). We have not directly tested social cognitive theories of other-race effects, which argue that people are not motivated to attend to and discriminate other-race faces, which consequently impairs their ability to recognise them (Rodin, 1987). However, Asian-Americans' higher accuracy for categorising other-race faces suggests that people can be motivated to attend to and learn how to discriminate other-race faces (see also Wan et al., 2015).

A considerable strength of the present studies was the use of faces that were deemed typical of their nation by individuals from those nations. Indeed, we selected nations that were deemed likely to include faces that look typical for those nations. Thus our results may not generalize to faces from all nations, particularly those that have been populated largely by immigration (either current or historic) from diverse parts of the world. Nevertheless, they demonstrate that when visually distinct populations do exist, people can categorize the national origins of own-race faces better than those of other-race faces, thus extending other-race effects to this new domain.

We have shown an advantage for categorising faces within one’s own race. An interesting extension would be to see if people are better at categorising own-race faces of their own nationality. One previous study has examined a related question. Coetzee, Greeff, Barrett, and Henzi (2009) asked black South Africans from several different ethnic groups to categorize black faces from two ethnic groups (Pedi, Tswana) by ethnic group. Accuracy for categorising own-ethnicity faces was no better than chance. However, it is not clear that the faces of Pedi and Tswana people are visually distinct, because they are part of the same larger Sotho ethnic group, they are geographically located in close proximity to each other and are highly intermixed (Hitzeroth, 1986). We have identified typical faces of six nations in Study 1, which are visually distinct from each other and could be used to examine own-nation face expertise through a national origins task.

The present results add to the current literature on other-race effects in face perception, which have been shown in recognition memory (Meissner & Brigham, 2001), age estimation (Dehon & Brédart, 2001), sex discrimination (O’Toole, Peterson, & Deffenbacher, 1996), gaze direction (Collova et al., 2017; Pavan, Dalmaso, Galfano, & Castelli, 2011), and making inferences about a persons’ mental state (Adams et al., 2010). The present results demonstrate a new other-race effect, through a categorisation task requiring subtle discrimination of faces from different nations. They also demonstrate that the effect is closely linked to experience with other-race faces and can be reversed with extensive other-race experience, consistent with the perceptual expertise account.

References

Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., … Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97-108.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology, 8*(4), 306-318.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42*(4), 673-681.

Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the "other-race" face classification advantage. *NeuroReport, 15*(5), 905-910.

Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual and Motor Skills, 64*(1), 319-325.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology, 48*(4), 879-894.

Coetzee, V., Greeff, J. M., Barrett, L., & Henzi, S. P. (2009). Facial-based ethnic recognition: Insights from two closely related but ethnically distinct groups. *South African Journal of Science, 105*, 464-466.

Collova, J. R., Kloth, N., Crookes, K., Burton, N., Chan, C.Y.H., Hsiao, J. H., & Rhodes, G. (2017). A new other race for gaze perception. *Journal of Experimental Psychology: Human perception and performance, 43*(11), 1857-1863

Corenblum, B., & Meissner, C. A. (2006). Recognition of faces of ingroup and outgroup children and adults. *Journal of Experimental Child Psychology, 93,* 187-206.

de Heering, A., de Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, *13*(1), 181-187.

Dehon, H., & Brédart, S. (2001). An ‘other-race’ effect in age estimation from faces. *Perception, 30*(9), 1107-1113.

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.

Feinman, S., & Entwisle, D. R. (1976). Children’s ability to recognize other children’s faces. *Child Development*, *47*(2), 506-510.

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., ... Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception, 38*(8), 1199-1210.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*, 201-210.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45-56.

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21*(9-10), 1224-1247.

Hitzeroth, H. W. (1986). On the genetic interrelationships of South African negroes. *American Journal of Physical Anthropology, 69,* 389-401.

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS One, 4*(1). e4215.

Lee, K., Anzures, G., Quinn, P. C., Pascalis, O., & Slater, A. M. (2011). Development of face processing in expertise. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 753-778). New York, NY: Oxford University Press.

Levin, D. T. (1996). “Classifying faces by race: The structure of face categories”. *Journal of Experimental Psychology: Learning, Memory, and Cognition 22*(6), 1364-1382.

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*(4), 559-574.

Lindsay, D. S., Jack, P, C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76*(4), 587-589.

Longmore, C. A., Liu, C. H., & Young, A. W. (2015). The importance of internal facial features in learning new faces. *The Quarterly Journal of Experimental Psychology, 68*(2), 249-260.

Malpass, R. S. & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330-334.

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., …, & O’Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test-Australian. *Cognitive Neuropsychology, 28*(2), 109–146

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy and Law, 7*(1), 3-35.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own- and other-race faces. *Visual Cognition, 21*(9-10), 1287-1305.

Ng, W., & Lindsay, R. C. L. (1994). Cross-race facial recognition. Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology 25*(2), 217-232.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). "Structural aspects of face recognition and the other-race effect" *Memory & Cognition, 22*, 208-224.

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

Pavan, G., Dalmaso, M., Galfano, G., & Castelli, L. (2011). Racial group membership is associated to gaze-mediated orienting in Italy. *PLoS One*, *6*(10).

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*, 291-303.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*. 313-331.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review, 13*(3), 499-505.

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition, 5*(2), 144-165.

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

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*(6), 440-444.

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in whites. *Basic and Applied Social Psychology, 22*(2), 71-84.

Sporer, S. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*, 36-97.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 18*(3), 315-327.

Tanaka, J. W., Heptonstall, B., & Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Visual Cognition, 21*(9-10), 1183-1201.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioural Neuroscience, 9*(1), 122-131.

Telzer, E. H., Flannery, J., Shapiro, M., Humphreys, K. L., Goff, B., Gabard-Durman, L., ... Tottenham, N. (2013). Early experience shapes amygdala sensitivity to race: An international adoption design. *The Journal of Neuroscience, 33*, 13484−13488.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. The *Quarterly Journal of Experimental Psychology Section A, 44*(4), 671-703.

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

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition, 144*, 91-115.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). "A field study of own-race advantage in South Africa and England". *Psychology, Public Policy, & Law, 7*(1), 119-133.

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review, 15*(6), 1093-1099.

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research, 105*, 61-69.

Supplemental Materials

Coherence analysis

Participants may have been able to detect subtle differences between faces of different nations and group them together, but assigned the wrong labels to these groups. To examine this possibility, we re-analysed our data, by looking at the proportion of faces of the same nation that participants grouped together with the same label, regardless of the correctness of that label. A three-way repeated measures analysis of variance (RM ANOVA) was conducted on these coherence values, with participant group (Caucasian, Asian, Asian-American) as a between participants variable, and race of face (own-race and other-race) and hair (with-hair, hair-masked) as within-participant variables. Planned t-tests were used to follow up interactions, where appropriate.

The 3-way ANOVA yielded significant main effects for face race, $F(1, 207) = 23.12$, $p < .001$, $\eta_p^2 = .10$, hair, $F(1, 207) = 11.67$, $p = .001$, $\eta_p^2 = .05$, and participant group, $F(2, 207) = 5.83$, $p = .003$, $\eta_p^2 = .05$. However, these effects were qualified by interactions between participant group and face race $F(2, 207) = 71.97$, $p < .001$, $\eta_p^2 = .41$, and face race and hair $F(1, 207) = 4.46$, $p = .036$, $\eta_p^2 = .02$, but not between hair and participant group $F(2, 207) = 0.19$, $p = .829$, $\eta_p^2 = .01$).

Caucasian participants showed higher coherence values for own-race faces (*M* = .20, *SD* = .10) than other-race faces (*M* = .06, *SD* = .03), $t(103) = 14.91$, $p < .001$, $d = 1.46$. Asian participants also showed higher coherence values for own-race faces (*M* = .20, *SD* = .14) than for other-race faces (*M* = .14, *SD* = .09), $t(53) = 3.48$, $p = .001$, $d = 0.47$. In contrast, Asian-American participants showed higher coherence values for other-race faces (*M* = .17, *SD* = 1.0) than own-race faces (*M* = .09, *SD* = .08), $t(51) = -5.58$, $p < .001$, $d = 0.77$.

All participants showed higher coherence values for own-race faces with-hair (*M* = .18, *SD* = .13), than other-race faces with hair (*M* = .11, *SD* = .09), $t(209) = 6.47$, $p < .001$, $d = .47$. All participants also showed higher coherence values for own-race faces with hair-masked (*M* = .16, SD = .11), than other-race faces with hair-masked (*M* = .11, *SD* = .09), $t(209) = 5.76$, $p < .001$, $d = .43$.