Neural dynamics of racial categorization predicts racial bias in face recognition and altruism

Yuqing Zhou [1,2,3], Tianyu Gao [1,2,3], Ting Zhang[1,2,3], Wenxin Li[1,2,3], Taoyu Wu[1,2,3], Xiaochun Han[1,2,3] and Shihui Han [1,2,3]*

The classification of individuals into different racial groups provides a precondition for racial bias in cognition and behaviour, but how the brain enables spontaneous racial categorization is not fully understood. Here using multimodal brain imaging measures, including electroencephalography, functional magnetic resonance imaging and magnetoencephalography, we probe the neural dynamics of racial categorization by quantifying the repetition suppression of neural responses to faces of different individuals of each racial group (that is, Asian, black or white). We show that categorization of other-race faces engages early two-stage dynamic activities in neural networks consisting of multiple interactive brain regions. Categorization of same-race faces, however, recruits a different and simple network in a later time window. Dynamic neural activities involved in racial categorization predict racial biases in face recognition and altruistic intention. These results suggest that there are distinct neural dynamics by which the brain sorts people into different racial groups as a social ground for cognition and action.

Despite the dominant cultural norms that oppose racial discrimination, persistent and pervasive biases against individuals due to their perceived race exist in employment and housing markets[1], clinical treatment[2], and jury judgments[3], among other situations. There is even an increasing trend in racial resentment in the twenty-first century[4]. To understand the psychological drive and the neural basis of racial discrimination shown in social behaviour, researchers have discovered racial biases in cognition and brain activity, such as poor recognition of other-race (OR) compared with same-race (SR) faces[5], implicit negative attitudes towards OR individuals[6,7] and decreased empathic brain responses to the suffering of OR individuals compared with SR individuals[8,9]. These racial biases are ubiquitous, as the racial ingroup favouritism in face recognition and empathic brain activity has been reported in different ethnic groups and in numerous societies[5,10]. The literature also shows that there is racial ingroup favouritism in face recognition[11], implicit attitude[12] and empathic brain activity[8,13,14] in Chinese society, where over 90% of the population are Han Chinese and there are few interracial interactions in real-life situations for the majority. The social demand of understanding of and intervening against racial biases is common across different societies.

To classify individuals into different racial groups, that is, racial categorization, at the cost of ignoring the identity of each individual, might have helped track alliances during evolution[15]. However, racial categorization provides a precondition for racial biases in behaviour, cognition and brain activity[10,16], which collide with egalitarian norms for a race-blind society[17]. Recent psychological models have assumed an increased motivation to view OR people categorically but to perceive SR people individually[18,19]. The processes of racial categorization are further proposed to be associated with racial biases in memory, emotion and social decision-making[10,18,19]. However, to date, how the human brain enables spontaneous categorization of some individuals to one racial group but others to a different racial group is not fully understood. Previous brain imaging studies of race focused on differences in neural responses to faces of two racial groups. For example, electroencephalography

(EEG) research found that participants, most of whom were white, showed increased amplitudes of event-related potentials (ERPs) to black compared with white faces, including an early negative activity (N1) peaking at about 100 ms after presentation of the face ('face onset'), and a following positive activity (P2) over the frontal region[20]. Differential neural responses to white versus black faces were also observed in the amplitude of subsequent negative activity (N2). These ERP effects have been observed in numerous studies regardless of whether experimental tasks required explicit attention to the racial identity of faces[21–25]. A similar approach has been adopted in functional magnetic resonance imaging (fMRI) studies, which compared blood-oxygen-level-dependent (BOLD) responses to black and white faces and have uncovered differential responses to black versus white faces (or vice versa) in several brain regions, including the amygdala[26,27], the fusiform gyrus[28], the anterior cingulate cortex[29], the posterior cingulate cortex (PCC)[30] and the dorsal lateral prefrontal cortex[31]. These neuroimaging findings unravelled differential neural responses to faces of two racial categories (that is, OR versus SR faces) and expanded our understanding of neural processes related to race.

However, the approach of focusing on differential neural responses to OR versus SR faces leaves open several issues regarding the neural mechanisms underlying racial categorization. First, the racial categorization of faces emphasizes group identity and relevant commonness shared by different individuals of each racial group. Racial categorization is mediated by neural processes that are common for different individuals to group them to a specific racial category. The approach of focusing on differential neural responses was limited to comparisons of brain activity to faces of two racial categories and did not allow researchers to disentangle neural processes to construct separate neural models of categorization of faces of each racial group. Second, the approach that examined differential neural responses was restricted by a same/other race dichotomy that does not fully depict real-life situations whereby people deal with multiple racial groups (for example, Asian, black and white). Behavioural research has provided evidence of differences among

¹School of Psychological and Cognitive Sciences, Peking University, Beijing, China. ²PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. ³Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China. *e-mail: shan@pku.edu.cn

multiple racial groups in perceived social status[32], stereotype (for example, threat)[33] and cognition (for example, face recognition)[34]. It is difficult to clarify whether the brain engages the same or different neural processes to sort perceived faces into multiple racial groups only by analysing differential brain responses to faces of two racial categories. Third, face perception consists of numerous processes[35], and differential neural responses to OR versus SR faces were observed in neural circuits underlying complicated and interactive processes of facial features and related emotion, attitude and stereotype[36]. Previous studies of race usually employed a single brain-imaging technique[23,36]. Little research has incorporated one paradigm with multimodal imaging measures to characterize spatiotemporal dynamics of multiple neural structures involved in the racial categorization of faces. Finally, although theoretical models have proposed associations between early racial categorization of OR faces and racial biases in memory, emotion and social decision-making[10,18,19], the underlying neural mechanism remains unclear.

The current work sought to combine multimodal brain imaging techniques, including EEG and ERPs, fMRI and magnetoencephalography (MEG), to determine the dynamic neural mechanisms underlying spontaneous racial categorization by quantifying repetition suppression (RS) of brain responses to faces of different individuals of each racial group (that is, Asian, black or white). RS reflects the attenuation in neural responses to a repeated occurrence of stimuli that engage common neuronal populations or common neural processes due to habituation[37]. RS of neural activity has been demonstrated in the firing rates of neurons measured using single-cell recording in monkeys[38] and in the activity of neuronal populations measured using neuroimaging techniques including EEG, MEG and fMRI[37]. Findings of RS of brain activity measured using EEG, MEG and fMRI allow researchers to identify mental and neural processes of a stimulus class in a specific time window or brain region. Neural underpinnings of categorization of faces of a specific racial group can be tested by examining decreased neural responses when the faces of individuals of that racial group are viewed repeatedly and the process of racial categorization is habituated. Previous research has examined RS of neural responses to estimate the impact of racial categorization on neural processes of individual identity and facial expression[14,39]. To our knowledge, previous research has not used multimodal neuroimaging measures to investigate the neural underpinnings of racial categorization of faces by inspecting RS of neural responses.

To disentangle the neural mechanisms of categorization of OR and SR faces, we modified a paradigm for studying RS of brain activity[40] by including an alternating condition (Alt-Cond), in which half of the SR faces and half of the OR faces of different individuals were presented alternately in a random order, and a repetition condition (Rep-Cond), in which only SR faces or OR faces of different individuals were displayed in a random order (Fig. 1a; Methods; Supplementary Fig. 1). Participants were asked to press a button to respond to a casual target face that was displayed in two consecutive trials. This one-back task did not require intentional racial categorization and thus allowed us to investigate the spontaneous neural coding of category-diagnostic features of different faces of a specific racial group. We hypothesized that if spontaneous racial categorization occurs, the neural activities supporting the grouping faces of different individuals into each racial class should be more attenuated in the Rep-Cond than in the Alt-Cond due to habituation (that is, the RS effect). The RS effect was quantified by comparing the neural responses to the same set of non-target faces of each racial group in the Rep-Cond versus the Alt-Cond. This allowed us to disentangle the neural processes of the categorization of faces of each racial group (for example, Asian, black and white) by controlling the processing of low-level perceptual features that may be different between faces of two racial groups. We integrated this paradigm with EEG and fMRI to segregate time courses and neural structures

engaged in the categorization of faces of each racial group. We also combined the same paradigm with MEG, which has the potential of millisecond source imaging[41], to validate our EEG and fMRI results and to scrutinize dynamic interactions of multiple nodes of neural circuits engaged in the racial categorization of faces during the first few hundreds of milliseconds after face onset.

In experiments 1–4, we recorded EEG images with millisecond time resolution to examine the time courses of the neural processes involved in the racial categorization of faces. The results unravelled distinct time courses of categorization of OR and SR faces and identified the key role of perceived racial relationships and facial configuration information in determining the neural processes underlying racial categorization. In experiment 5, we employed fMRI to record BOLD signals with millimetre spatial resolution to examine the neural networks underlying racial categorization. The results revealed distinct neural networks involved in the categorization of OR and SR faces and of different subgroup of OR faces. In experiment 6, we recorded MEG images to further analyse the spatiotemporal dynamics of the neural networks involved in the categorization of faces from each racial group and examined possible associations between the neural underpinnings of racial categorization and racial biases in cognition and emotion. The results suggested that there are distinct dynamic neural models of categorization of faces from different racial groups, which further predict racial biases in face recognition, empathy and altruistic intentions.

Results

Pilot experiment: distinct perceived categories of Asian, black and white faces. Because we were using Asian, black and white faces in our multimodal imaging experiments in different samples, we sought to verify the subjective racial categories of these face stimuli in a pilot experiment. To this end, we asked Chinese adults ($N = 57$) to perform a similarity rating task and report their perceived similarity of two simultaneously presented faces on a 9-point scale, where 1 represents not similar at all and 9 represents extremely similar (see Methods for details). The mean similarity matrix across all participants was subjected to the ALSCAL multidimensional scaling procedure, which resulted in a four-dimensional solution for both male and female faces (stress = 0.052 and 0.066, respectively). The matrix of mean similarity scores was collapsed for male and female faces across all participants to estimate perceived clusters of faces in a face space (Fig. 1b). Independent sample t-tests showed that the mean distance between each face and other SR faces in the face space was significantly shorter than that between each face and OR faces (Asian versus non-Asian faces: $t(30) = -23.98$, $P < 0.001$, Cohen's $d = 8.48$, 95% confidence interval (CI) = -2.28, -1.91; black versus non-black faces: $t(30) = -88.76$, $P < 0.001$, Cohen's $d = 31.38$, 95% CI = -3.00, -2.86; white versus non-white faces: $t(30) = -20.00$, $P < 0.001$, Cohen's $d = 7.07$, 95% CI = -2.09, -1.70). These results indicate that the Asian, black and white faces used in our work were indeed perceived in different racial categories. Moreover, black faces ($M = 0.51$) were more densely clustered than white ($M = 1.38$) and Asian ($M = 1.25$) faces ($F(2,45) = 45.59$, $P < 0.001$, $\eta_P^2 = 0.670$, 90% CI = 0.51, 0.74; post hoc analyses (Bonferroni-corrected): black versus Asian: $P < 0.001$, 95% CI = -0.99, -0.50; black versus white: $P < 0.001$, 95% CI = -1.12, -0.63; Asian versus white: $P = 0.601$, 95% CI = -0.12, 0.37), which suggests that there were greater feelings of similarity among black faces than white/Asian faces. Furthermore, the distance between Asian and black clusters ($M = 3.51$) was longer than that between Asian and white clusters ($M = 3.18$) ($t(30) = 6.63$, $P < 0.001$, Cohen's $d = 2.34$, 95% CI = 0.23, 0.43) in the face space. These results implicate possible distinct processes involved in the categorization of faces from different racial groups, which were tested in the following neuroimaging experiments.
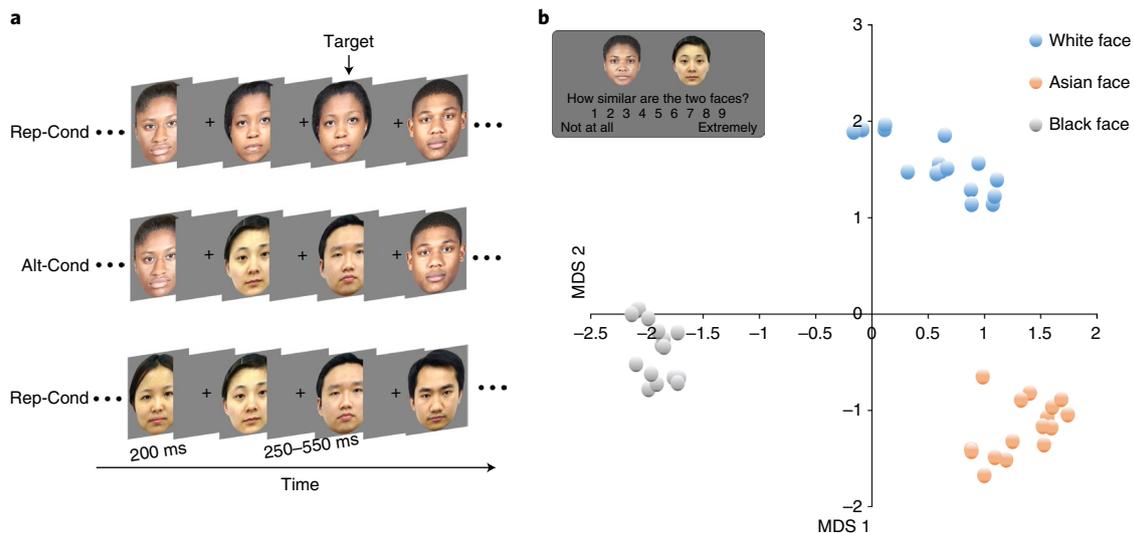
Fig. 1 | Stimuli and behavioural results. a, Illustration of the stimuli and procedures in the Rep-Cond and the Alt-Cond in experiment 1a. There were equal numbers of target faces in the Alt-Cond and the Rep-Cond. **b**, Stimuli, procedures and results of Chinese participants in the pilot experiment. Asian, black and white faces were clustered in a two-dimensional face-space built on the basis of similarity rating scores in the pilot experiment ($N = 55$). Black faces ($M = 0.51$) were more densely clustered than white ($M = 1.38$) and Asian ($M = 1.25$) faces. $F(2,45) = 45.59$, $P < 0.001$, $\eta_P^2 = 0.670$, 90% CI = 0.51, 0.74; post hoc analyses. Bonferroni-corrected: black versus Asian: $P < 0.001$, CI = −0.99, −0.50; black versus white: $P < 0.001$, CI = −1.12, −0.63; Asian versus white: $P = 0.601$, CI = −0.12, 0.37. The distance between Asian and black clusters ($M = 3.51$) was longer than that between Asian and white clusters ($M = 3.18$) ($t(30) = 6.63$, $P < 0.001$, Cohen's $d = 2.34$, CI = 0.23, 0.43) in the face space. MDS, multidimensional space. Face images reproduced from refs. [13] and [75].

Experiment 1a: distinct time courses of neural categorization of OR and SR faces. In experiment 1a we tested the hypothesis that there is an earlier categorization of OR faces than of SR faces, which is based on psychological models of race perception[10,18,19]. We performed EEG recordings to estimate time courses of RS effects on neural responses to OR faces and SR faces, respectively. This hypothesis predicts that RS of neural responses occurs earlier for OR faces than for SR faces. Chinese students ($N = 38$) performed the one-back task on Asian, black and white faces in Rep-Cond and Alt-Cond (duration of 200 ms, interstimulus interval of 250–550 ms) during EEG recording. Asian/black and Asian/white faces were presented in different EEG sessions to examine RS of neural responses to subcategories of OR faces (black and white faces) and to SR faces (Asian faces) (see Methods), respectively. Participants detected targets (10% of all trials) with high response accuracies (>80% in all conditions; Supplementary Table 1). The repeated-measures analysis of variance (ANOVA) of reaction times and response accuracies with race (Asian versus black or Asian versus white) × condition (Alt-Cond versus Rep-Cond) as within-subject variables did not show any significant effect with respect to race (see Supplementary Table 1 for statistical details), thus showing that there is no evidence for differences in attentional demand and task difficulty between the detection of OR and SR targets.

Non-target faces elicited negative activity at 95–125 ms (N1) followed by positive activity at 140–200 ms (P2) and negative activity at 200–400 ms (N2) over the frontal/central electrodes (Fig. 2). The mean amplitude of each component at these electrodes was subject to ANOVAs with race and condition as within-subject variables to examine RS effects (that is, decreased neural responses in Rep-Cond versus Alt-Cond) in different time windows of neural responses to the faces of each racial group. There was a significant main effect of race on N1 amplitudes in Asian/black sessions ($F(1,37) = 6.81$, $P = 0.013$, $\eta_P^2 = 0.155$, 90% CI = 0.02, 0.32) but not in Asian/white sessions ($F(1,37) = 2.50$, $P = 0.122$, $\eta_P^2 = 0.063$, 90% CI = 0, 0.21), thus suggesting that there was a larger N1 amplitude to black than Asian faces. This is consistent with previous findings of a larger N1 amplitude to black than white faces in white samples[20].

However, there was not a significant difference in the N1 amplitudes between the Alt-Cond and the Rep-Cond (Asian/black sessions: $F(1,37) = 0.46$, $P = 0.500$, $\eta_P^2 = 0.012$, 90% CI = 0, 0.12; Asian/white sessions: $F(1,37) = 1.25$, $P = 0.271$, $\eta_P^2 = 0.033$, 90% CI = 0, 0.16). To complement the ANOVA results, we applied a Bayes factor approach to assess the difference in the N1 amplitudes between the Alt-Cond and the Rep-Cond. The Bayes factor of comparing the model containing the main effect of condition to the null model was also in agreement with the absence of RS of the N1 amplitudes (Bayes factor = 0.21 and 0.30, respectively). Together, these results show that there is no evidence for RS of the N1 amplitude, even though the N1 differed significantly between black and Asian faces.

ANOVAs of the P2 amplitudes (140–200 ms) first revealed significant effects of condition for both Asian/black and Asian/white EEG sessions ($F(1,37) = 21.17$ and 5.96, $P < 0.001$ and $P = 0.02$, $\eta_P^2 = 0.364$ and 0.139, 90% CI = (0.16, 0.52) and (0.01, 0.30), respectively), thus indicating that there were reliable decreased P2 amplitudes in the Rep-Cond than in the Alt-Cond. Moreover, ANOVAs of the P2 amplitudes showed significant interactions of race × condition for both Asian/black and Asian/white EEG sessions ($F(1,37) = 19.85$ and 9.74, $P < 0.001$ and $P = 0.003$, $\eta_P^2 = 0.349$ and 0.208, 90% CI = (0.15, 0.50) and (0.04, 0.38), respectively; see Extended Data Fig. 1 for statistical details). Simple effect analyses confirmed RS of P2 amplitudes to black and white faces ($F(1,37) = 31.56$ and 13.50, $P < 0.001$ and $P = 0.001$, $\eta_P^2 = 0.460$ and 0.267, 90% CI = (0.25, 0.59) and (0.08, 0.43), respectively), but not to Asian faces (Asian/black and Asian/white sessions: $F(1,37) = 0.32$ and 0.012, $P = 0.576$ and 0.912, $\eta_P^2 = 0.009$ and 0.000, 90% CI = (0, 0.11) and (0, 0.03), respectively), thus indicating that there are significant RS effects in response to OR faces but not to SR faces. Additional Bayes factor analyses also suggested that there was an absence of RS of the P2 amplitudes to Asian faces in both Asian/black and Asian/white sessions when comparing the null hypothesis over the alternative model (Bayes factor = 0.20 and 0.17, respectively). The voltage topographies showed that the maximum RS effect on P2 amplitudes was over the frontal/central regions for both black and white faces (Fig. 2).
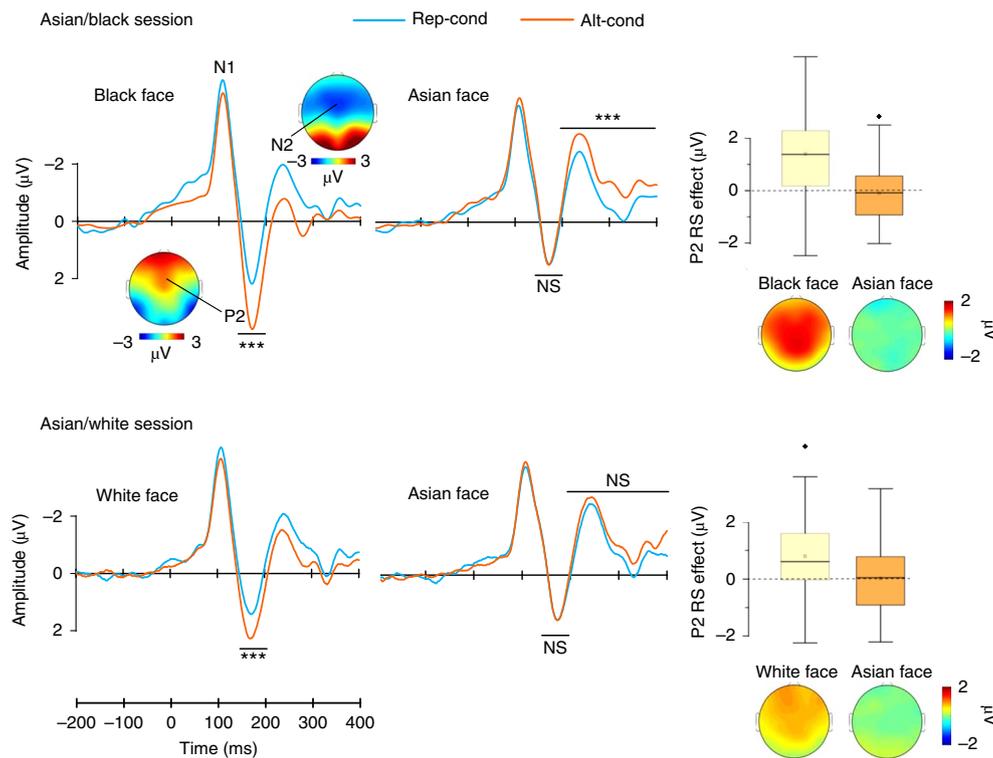
Fig. 2 | ERP results of Chinese participants in experiment 1a. Left: ERPs to OR faces and SR faces at frontal/central electrodes (the *x* axis at the bottom of the figure is applicable to all the ERP charts). Right: box plots illustrating the RS effects (that is, decreased amplitudes in the Rep-Cond compared with the Alt-Cond) on P2 amplitudes to OR and SR faces. The plots show the quartiles (boxes), means (square inside boxes), medians (horizontal lines inside boxes), maximum and minimum excluding outliers (whiskers), and outliers (data points higher or lower than 1.5-times the quartile (25th and 75th percentile) values) (diamonds). The voltage topographies show locations of the maximum P2 and N2 amplitudes and maximum RS effects on P2 amplitudes over the frontal/central regions for both black and white faces. There was significant RS of P2 amplitudes to black and white faces ($N=38$, $F(1,37)=31.56$ and 13.50, $P<0.001$ and $P=0.001$, $\eta_P^2=0.460$ and 0.267, 90% CI = (0.25, 0.59) and (0.08, 0.43), respectively), but not to Asian faces (Asian/black and Asian/white sessions: $F(1,37)=0.32$ and 0.012, $P=0.576$ and $P=0.912$, $\eta_P^2=0.009$ and 0.000, 90% CI = (0, 0.11) and (0, 0.03), respectively). There was significant RS of N2 amplitudes to Asian faces in Asian/black sessions ($F(1,37)=18.56$, $P<0.001$, $\eta_P^2=0.334$, 90% CI = 0.13, 0.49) but not in Asian/white sessions ($F(1,37)=1.98$, $P=0.167$, $\eta_P^2=0.051$, 90% CI = 0, 0.19). NS, not significant, ***$P\leq0.001$.

There were also significant race×condition interactions on N2 amplitudes (200–400 ms) for both Asian/black and Asian/white EEG sessions ($F(1,37)=26.44$ and 8.86, $P<0.001$ and $P=0.005$, $\eta_P^2=0.417$ and 0.193, 90% CI = (0.21, 0.56) and (0.04, 0.36), respectively). Separate analyses confirmed RS of N2 amplitudes to Asian faces in Asian/black sessions ($F(1,37)=18.56$, $P<0.001$, $\eta_P^2=0.334$, 90% CI = 0.13, 0.49), but not in Asian/white sessions ($F(1,37)=1.98$, $P=0.167$, $\eta_P^2=0.051$, 90% CI = 0, 0.19), which suggests that there were decreased N2 amplitudes in the responses to SR faces when being presented alternately with black faces than when being presented repeatedly. Additional Bayes factor analyses also suggested that there was an absence of RS of N2 amplitudes to Asian faces in Asian/white sessions when comparing the null hypothesis over the alternative model (Bayes factor = 0.44). Together, the P2 and N2 results support our hypotheses and showed evidence for early categorization processes of OR faces in the P2 time window and late categorization processes of SR faces in the N2 time window. Moreover, the N2 RS effect seemed to depend on the context of the presence of black faces.

According to the psychological models that assume a parallel occurrence of enhanced categorization and decreased individuation processes of faces[10,18,19], one would expect that a weakened categorization of OR faces in the Rep-Cond (versus the Alt-Cond), as reflected in P2 RS effects, would be associated with enhanced individuation processing of OR faces. To test this prediction, we calculated the mean amplitude of the N170 component (140–200 ms)

over the lateral occipital/temporal electrodes, which is considered to be sensitive to face identity[39,40], in the Rep-Cond and the Alt-Cond. We then assessed correlations between the differential N170 amplitudes as an index of individuation processing and differential P2 amplitudes as an index of categorization processing in the Rep-Cond versus the Alt-Cond. The results revealed that a larger P2 RS effect (that is, decreased P2 amplitudes in the Rep-Cond versus the Alt-Cond) predicted greater N170 enhancement (that is, increased N170 amplitudes in the Rep-Cond versus the Alt-Cond) for both black and white faces ($r=0.512$ and 0.520, $P=0.001$ and $P=0.001$, 95% CI = (0.25, 0.71) and (0.22, 0.74), respectively; Supplementary Fig. 2). These results further support the functional role of neural responses in the P2 time window in the categorization of OR faces and suggest a neural basis of the reversal between categorization processes and individuation processes of OR faces. In the following EEG experiments, we focused on P2 and N2 RS effects, which reflect the main processes relevant to the spontaneous racial categorization of faces.

Experiment 1b: replication of distinct time courses of SR/OR categorization. Are the distinct RS effects on neural responses to OR and SR faces independent of a specific set of face stimuli and testing samples? To clarify this, in experiment 1b, we recorded EEG traces from an independent Chinese sample ($N=38$) who performed the one-back task on a different set of Asian, black and white faces cropped to remove hair (experiment procedures were the

same as those for experiment 1a, see Methods for details of stimuli and participants, and see Supplementary Results 1, Supplementary Fig. 3 and Supplementary Table 2 for behavioural results).

ANOVAs of P2 amplitudes in both Asian/black and Asian/white sessions showed significant interactions of race×condition ($F(1,37)=39.83$ and 9.86, $P<0.001$ and $P=0.003$, $\eta_P^2=0.518$ and 0.210, 90% CI = $(0.32, 0.64)$ and $(0.05, 0.38)$, respectively; see Extended Data Fig. 2 for statistical details). Simple effect analyses confirmed significantly decreased P2 amplitudes in the Rep-Cond than in the Alt-Cond for black and white faces ($F(1,37)=38.74$ and 14.09, $P<0.001$ and $P=0.001$, $\eta_P^2=0.511$ and 0.276, 90% CI = $(0.31, 0.63)$ and $(0.09, 0.44)$, respectively) but not for Asian faces (Asian/black and Asian/white sessions: $F(1,37)=1.12$ and 0.007, $P=0.296$ and $P=0.932$, $\eta_P^2=0.029$ and 0.001, 90% CI = $(0, 0.16)$ and $(0, 0.01)$, Bayes factor = 0.29 and 0.18, respectively; Supplementary Fig. 4). ANOVAs of the N2 amplitude (200–270 ms) also showed significant race×condition interactions for both Asian/black and Asian/white EEG sessions ($F(1,37)=50.47$ and 16.57, $P<0.001$ and $P<0.001$, $\eta_P^2=0.577$ and 0.309, 90% CI = $(0.38, 0.68)$ and $(0.11, 0.47)$, respectively). Simple effect analyses further verified that the N2 amplitudes to Asian faces were significantly decreased in the Rep-Cond than in the Alt-Cond in Asian/black sessions ($F(1,37)=6.04$, $P=0.019$, $\eta_P^2=0.140$, 90% CI = $0.01, 0.31$) but not in Asian/white sessions ($F(1,37)=2.20$, $P=0.147$, $\eta_P^2=0.056$, 90% CI = $0.01, 0.20$, Bayes factor = 0.48). Similarly, there was a significant positive correlation between the N170 enhancement effect and the P2 RS effect for both black faces ($r=0.566$, $P<0.001$, 95% CI = $0.29, 0.77$) and white faces ($r=0.443$, $P=0.005$, 95% CI = $0.13, 0.70$) across all participants (Supplementary Fig. 5). These results replicated the findings of early P2 RS effects on neural responses to OR faces and late context-dependent N2 RS effects on neural responses to SR faces in an independent sample and using a different set of facial stimuli.

Experiment 2: distinct time courses of SR/OR categorization with controlled gender categorization. Categorization of faces along different dimensions such as race and gender are assumed to take place in parallel[42], and differential neural responses to male and female faces are also observed in the P2 time window[20]. Because both male and female faces were used in both the Alt-Cond and the Rep-Cond in experiments 1a and 1b, it is unclear to what degree the categorization of faces in terms of gender might have affected the early P2 RS effect related to OR faces and the late N2 RS effect related to SR faces observed in these experiments. In experiment 2, we addressed this issue by adopting a design that reduced the effect of gender categorization to a minimum. The experimental procedures in experiment 2 were the same as those for experiment 1a, except that faces of only one gender (either male or female) were presented in the Alt-Cond and the Rep-Cond (see Methods for details). Because faces of the same gender were presented in each condition, this design allowed us to separately assess the racial categorization of male and female faces when the gender categorization of faces was minimized in both the Alt-Cond and the Rep-Cond due to habituation. We recorded EEG traces from an independent Chinese sample ($N=36$) in Asian/black sessions and another independent Chinese sample ($N=31$) in Asian/white sessions. This design simplified the context of OR faces in each sample (that is, viewing faces of only one subcategory of OR faces) and reduced EEG recording times. RS effects were examined separately for male and female faces in both Asian/black and Asian/white sessions, and the number of trials in each condition was the same as that in experiments 1a and 1b (see Supplementary Table 3 for behavioural results).

To compare RS effects on brain activities underlying racial categorization of faces between male and female faces, the P2 and N2 amplitudes to non-target faces were subjected to ANOVAs with race (Asian versus black or Asian versus white), condition (Alt-Cond versus Rep-Cond) and face gender (male versus female) as within-subject variables. Similarly, ANOVAs of P2 amplitudes showed significant race×condition interactions for both Asian/black sessions ($F(1,35)=23.45$, $P<0.001$, $\eta_P^2=0.401$, 90% CI = 0.19, 0.55) and Asian/white sessions ($F(1,30)=7.953$, $P=0.008$, $\eta_P^2=0.210$, 90% CI = $0.03, 0.39$). Simple effect analyses indicated significant RS of P2 amplitudes to black ($F(1,35)=56.91$, $P<0.001$, $\eta_P^2=0.619$, 90% CI = $0.43, 0.72$) and white faces ($F(1,30)=23.47$ $P<0.001$, $\eta_P^2=0.439$, 90% CI = $0.21, 0.59$), but not to Asian faces (Asian/black sessions: $F(1,35)=2.75$, $P=0.106$, $\eta_P^2=0.073$, 90% CI = $0, 0.23$, Bayes factor = 0.61; Asian/white sessions: $F(1,30)=2.29$, $P=0.141$, $\eta_P^2=0.071$, 90% CI = $0, 0.24$, Bayes factor = 0.24; Fig. 3). Importantly, the three-way interaction of race×condition×face gender was not significant (Asian/black: $F(1,35)=2.70$, $P=0.109$, $\eta_P^2=0.072$, 90% CI = 0, 0.23, Bayes factor = 0.33; Asian/white: $F(1,30)=2.88$, $P=0.100$, $\eta_P^2=0.088$, 90% CI = $0, 0.26$, Bayes factor = 0.40; Supplementary Fig. 6, see Extended Data Fig. 3 for statistical details), thus showing that there is no evidence for reliable differences in the patterns of RS of P2 amplitudes between male and female faces.

Similarly, ANOVAs of the N2 amplitudes also showed significant race×condition interactions for Asian/black sessions ($F(1,35)=23.69$, $P<0.001$, $\eta_P^2=0.404$, 90% CI = 0.19, 0.55) and Asian/white sessions ($F(1,30)=16.37$, $P<0.001$, $\eta_P^2=0353$, 90% CI = $0.13, 0.52$). Simple effect analyses revealed significant RS effects on the N2 amplitude to Asian faces in Asian/black sessions ($F(1,35)=12.32$, $P=0.001$, $\eta_P^2=0.260$, 90% CI = $0.07, 0.43$), but not in Asian/white sessions ($F(1,30)=1.94$, $P=0.174$, $\eta_P^2=0.061$, 90% CI = $0, 0.23$, Bayes factor = 0.35; Fig. 3). Similarly, the three-way interaction of race×condition×face gender was not significant (Asian/black: $F(1,35)=0.55$, $P=0.462$, $\eta_P^2=0.016$, 90% CI = 0, 0.13, Bayes factor = 0.28; Asian/white: $F(1,30)=1.90$, $P=0.178$, $\eta_P^2=0.060$, 90% CI = $0, 0.22$, Bayes factor = 0.54; see Extended Data Fig. 3 for statistical details), thus showing that there is no evidence for any significant effect of face gender on RS of N2 amplitudes to SR faces. These results verified the early P2 RS effect related to OR faces and the late context-dependent N2 RS effect related to SR faces when spontaneous gender categorization was controlled.

Experiment 3: racial relationship and neural processes involved in racial categorization. An important issue arising from the results obtained in experiments 1 and 2 is whether the RS of P2 and N2 amplitudes with respect to racial categorization manifested as a pure effect of racial identity of perceived faces or was determined by racial relationships between observers and perceived faces. To clarify these aspects, in experiment 3, we recorded EEG traces from an independent sample of white students ($N=35$). The stimuli and procedure in experiment 3 were the same as those for experiment 1a, except that white/black faces and white/Asian faces were presented in different EEG sessions. If the RS of P2 and N2 amplitudes observed in experiments 1 and 2 reflected modulations of neural activities due to racial relationships between observers and perceived faces, white participants should show RS effects on P2 amplitudes to black and Asian (that is, OR) faces, but RS effects on N2 amplitudes to white (that is, SR) faces (see Supplementary Table 4 for behavioural results).

As expected, ANOVAs of P2 amplitudes showed significant interactions of race×condition for both white/black sessions ($F(1,31)=31.56$, $P<0.001$, $\eta_P^2=0.504$, 90% CI = $0.28, 0.64$; inadequate EEG recordings were obtained for 3 participants in the white/black sessions) and white/Asian sessions ($F(1,34)=4.383$, $P=0.044$, $\eta_P^2=0.114$, 90% CI = $0, 0.28$; see Extended Data Fig. 4 for statistical details). Reliable RS of P2 amplitudes was observed for black ($F(1,31)=32.23$, $P<0.001$, $\eta_P^2=0.510$, 90% CI = $0.28, 0.64$) and Asian faces ($F(1,34)=9.81$, $P=0.004$, $\eta_P^2=0.224$, 90% CI = $0.05, 0.40$), but not for white faces (white/black sessions: ($F(1,31)=3.37$, $P=0.076$, $\eta_P^2=0.098$, 90% CI = $0, 0.27$, Bayes factor = 0.84; white/Asian
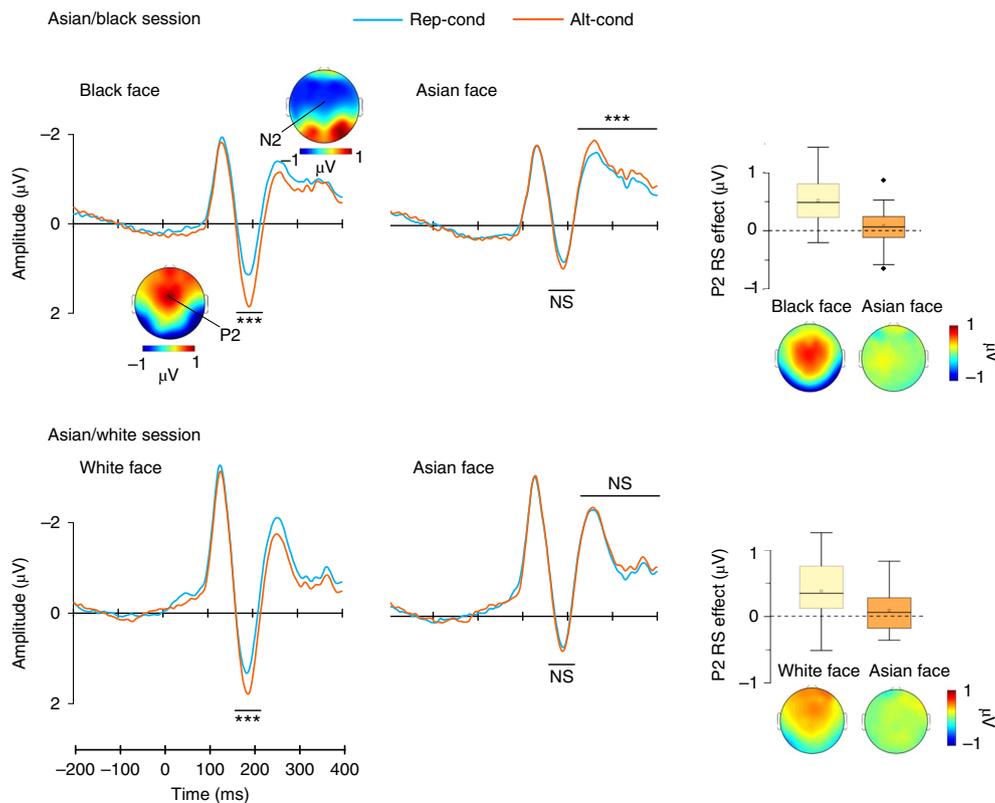
Fig. 3 | Results of Chinese participants in experiment 2. Left: ERPs to OR (black and white) and SR (Asian) faces at frontal/central electrodes (the *x* axis at the bottom of the figure is applicable to all the ERP charts). Right: box plots illustrating the RS effects on P2 amplitudes (that is, the difference in P2 amplitudes between the Alt-Cond and the Rep-Cond) to OR faces and SR faces. The plots show the quartiles (boxes), means (square inside boxes), medians (horizontal lines inside boxes), maximum and minimum excluding outliers (whiskers), and outliers (data points higher or lower than 1.5-times the quartile (25th and 75th percentile) values) (diamonds). The voltage topographies show locations of the maximum P2 and N2 amplitudes and the maximum RS effect on P2 amplitudes over the frontal/central regions for both black and white faces. There was significant RS of the P2 amplitudes to black faces ($N=36$, $F(1,35)=56.91$, $P<0.001$, $\eta_P^2=0.619$, 90% CI=0.43, 0.72) and white faces ($N=31$, $F(1,30)=23.47$ $P<0.001$, $\eta_P^2=0.439$, 90% CI=0.21, 0.59), but not to Asian faces (Asian/black sessions: $F(1,35)=2.75$, $P=0.106$, $\eta_P^2=0.073$, 90% CI=0, 0.23; Asian/white sessions: $F(1,30)=2.29$, $P=0.141$, $\eta_P^2=0.071$,90% CI=0, 0.24. There was significant RS of the N2 amplitude to Asian faces in Asian/black sessions ($F(1,35)=12.32$, $P=0.001$, $\eta_P^2=0.260$, 90% CI=0.07, 0.43) but not in Asian/white sessions ($F(1,30)=1.94$, $P=0.174$, $\eta_P^2=0.061$, 90% CI=0, 0.23). ***$P\leq0.001$.

sessions: $F(1,34)=0.31$, $P=0.582$, $\eta_P^2=0.009$, 90% CI=0, 0.12, Bayes factor=0.21; Fig. 4). Similarly, ANOVAs of N2 amplitude showed significant interactions of race × condition for white/black sessions ($F(1,31)=30.30$, $P<0.001$, $\eta_P^2=0.494$, 90% CI=0.27, 0.63) and white/Asian sessions ($F(1,34)=6.79$, $P=0.014$, $\eta_P^2=0.166$, 90% CI=0.02, 0.34). Simple effect analyses confirmed RS of N2 amplitudes to white faces in white/black sessions ($F(1,31)=19.40$, $P<0.001$, $\eta_P^2=0.385$, 90% CI=0.16, 0.54), but not in white/Asian sessions ($F(1,34)=2.69$, $P=0.110$, $\eta_P^2=0.073$, 90% CI=0, 0.23, Bayes factor=0.61; Fig. 4). These results parallel those in experiments 1 and 2 and further demonstrate the early P2 RS effect related to OR faces and the late N2 RS effect related to SR faces.

Overall, the results of experiments 1–3 uncovered distinct time courses of neural processes involved in the categorization of OR faces (140–200 ms after face onset) and SR faces (after 200 ms), which depended on racial relationships between observers and perceived faces. These effects were similarly observed in Chinese and white participants.

The results of experiments 1–3 also raise an interesting issue; that is, why was the RS effect on N2 amplitudes to SR faces evident when SR faces were presented alternately with black faces but not with non-black OR faces? Behavioural research has shown evidence of a stereotypical association between black faces and threats[43–45]. Moreover, perceived social threats may induce an increased sense of ethnic identity[46]. Based on these findings, we assumed that the

context-dependent N2 RS effect might reflect enhanced categorization of SR faces due to the presence of black faces. To test this account, we conducted an additional behavioural experiment to examine reaction times to SR faces presented alternately with black faces or non-black faces in an explicit racial categorization task. Twenty-seven Chinese students (14 males, mean age ± s.d. = 22.6 ± 2.2 years) and 20 white students (10 males, mean age ± s.d. = 22.7 ± 1.8 years) were recruited to this task. Participants were presented with faces used in experiment 1a (face duration of 200 ms, followed by a fixation cross of 1,200 ms) and asked to categorize faces of different individuals into two racial groups as fast and as accurately as possible. There were two sessions (2 blocks of 32 trials in each session) for both Chinese (Asian/black sessions plus Asian/white sessions) and white participants (white/black sessions plus white/Asian sessions). Our account predicted that categorization of SR faces would be faster when SR faces presented alternately with black than non-black OR faces. As expected, the results showed that for both Chinese and white participants, reaction times to SR faces were shorter when presented alternately with black faces compared with non-black OR faces (Chinese participants: 450 versus 502 ms, $t(26)=-5.791$, $P<0.001$, Cohen's $d=1.11$, 95% CI=−69.77, −33.22; white participants: 416 versus 451 ms, $t(19)=-3.166$, $P=0.005$, Cohen's $d=0.71$, 95% CI=−59.06, −12.05). These behavioural results support the account that black faces provide a context that facilitates neural processes involved in the categorization of SR faces.
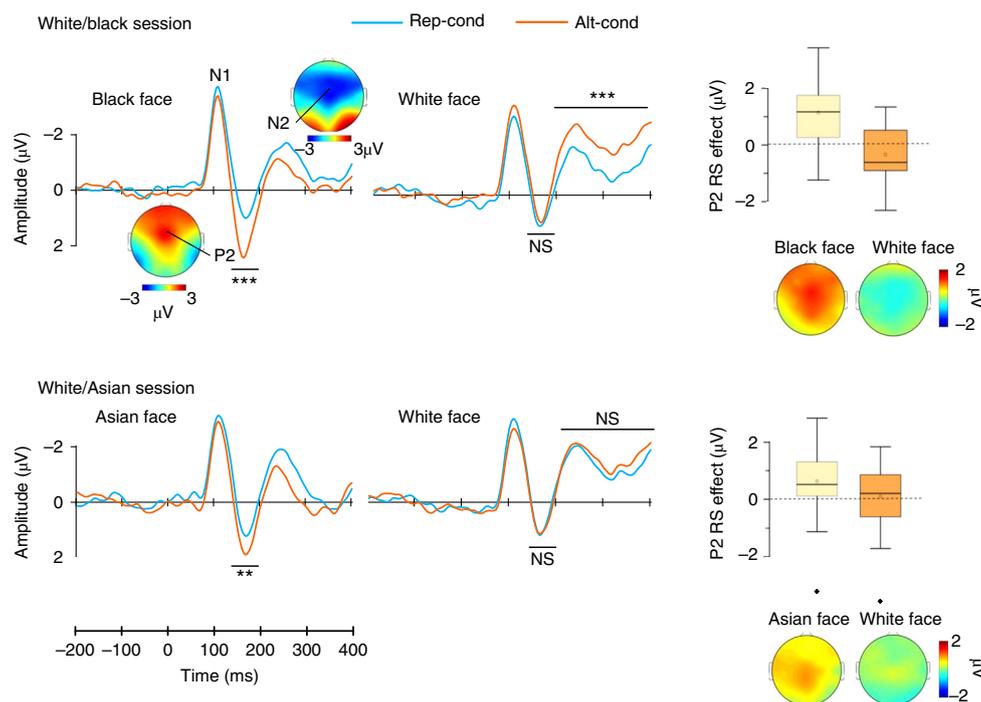
Fig. 4 | ERP results of white participants in experiment 3. Left: ERPs to OR faces and SR faces at frontal/central electrodes (the x axis at the bottom of the figure is applicable to all the ERP charts). Right: box plots illustrating the RS effects on P2 amplitudes to OR and SR faces. Plots show the quartiles (boxes), means (square inside boxes), medians (horizontal lines inside boxes), maximum and minimum excluding outliers (whiskers), and outliers (diamonds). The voltage topographies show locations of the maximum P2 and N2 amplitudes and the maximum RS effect on P2 amplitudes over the frontal/central regions for both black and Asian faces. There was significant RS of the P2 amplitudes to black faces ($N=32$, $F(1,31)=32.23$, $P<0.001$, $\eta_P^2=0.510$, 90% CI $=0.28$, 0.64) and Asian faces ($N=35$, $F(1,34)=9.81$, $P=0.004$, $\eta_P^2=0.224$, 90% CI $=0.05$, 0.40), but not to white faces (white/black sessions: ($F(1,31)=3.37$, $P=0.076$, $\eta_P^2=0.098$, 90% CI $=0$, 0.27; white/Asian sessions: $F(1,34)=0.31$, $P=0.582$, $\eta_P^2=0.009$, 90% CI $=0$, 0.12). There was significant RS of the N2 amplitude to white faces in white/black sessions ($F(1,31)=19.40$, $P<0.001$, $\eta_P^2=0.385$, 90% CI $=0.16$, 0.54), but not in white/Asian sessions ($F(1,34)=2.69$, $P=0.110$, $\eta_P^2=0.073$, 90% CI $=0$, 0.23). $**P<0.01$; $***P<0.001$.

Experiment 4: facial configuration and neural processes involved in racial categorization. Skin tone and facial configuration have been proposed to be two observable racial attributes shared by faces of a racial group[47,48]. In experiment 4, we sought to clarify whether RS of P2 amplitudes to OR faces and RS of N2 amplitudes to SR faces manifested the processing of skin tone or facial configuration during racial categorization. The stimuli and procedures in experiment 4 were the same as those for experiment 1a, except that faces were presented upside down. Face inversion does not change the skin tone of faces but disrupts facial configural information[49]. If skin tone is sufficient for the racial categorization of faces, we would expect similar P2 and N2 RS effects as those in experiments 1–3 when viewing inverted faces. However, if facial configuration is necessary for the racial categorization of faces, we would expect significantly decreased P2 and N2 RS effects for inverted faces.

We recorded EEG traces from an independent Chinese sample ($N=35$). As expected, response accuracies during the one-back task were relatively low (>62% in all conditions) due to perceived inverted faces (see Supplementary Table 5 for behavioural results). Moreover, the P2 amplitude did not show greater RS effects for inverted OR faces than SR faces. Similarly, the N2 amplitude did not show reliable RS effects for inverted SR faces (Supplementary Fig. 7, see Extended Data Fig. 5 for statistical details). To estimate whether the null effect was due to the sample sizes used for this experiment, we assessed how likely it was to observe a significant RS effect given the sample sizes in experiment 4 and the effect size observed in experiments 1–3 using G*Power[50]. The results showed that given the smallest effect size of the RS of P2 and N2 amplitudes observed across experiments 1–3 (that is, $\eta_P^2=0.114$), the sample

sizes in experiment 4 had a power of 0.86 to reveal reliable RS of P2 and N2 amplitudes, which is well above conventional recommendations[51]. The results of these analyses suggest that the null effect in experiment 4 was not simply due to underpowered sample sizes and reflected a possible disruption of P2 and N2 RS effects due to face inversion. In summary, because face inversion interferes with facial configural information but keeps skin tone unchanged[49], the results in experiment 4 implicate that skin tone is not sufficient, whereas facial configural information is necessary, for neural processes underlying distinct time courses of the categorization of OR and SR faces.

Experiments 5a and 5b: distinct neural architectures for the categorization of OR and SR faces. Because our ERP results demonstrated distinct time courses of neural processes involved in the categorization of OR and SR faces, in experiment 5a, we further investigated whether these neural processes recruit distinct neural networks by localizing RS effects on neural responses to OR faces and SR faces. In experiment 5a, we recorded BOLD signals, using fMRI, from an independent Chinese sample ($N=55$; 2 participants were excluded from analyses due to excessive head movement). Similar to the experiments above, participants performed the one-back task on Asian, black and white faces presented in the Alt-Cond and the Rep-Cond, but longer durations (400 ms) and interstimulus intervals (2, 4 or 6 s) were used to disentangle BOLD responses to each face (see Methods). A large sample was recruited because the long interstimulus intervals used might weaken the RS effects on the BOLD responses to faces (see Supplementary Table 6 for behavioural results).

As our EEG results showed similar time courses of the P2 RS effect for black and white faces, we first combined the BOLD responses to non-target black and white faces in the Alt-Cond and the Rep-Cond to examine RS of brain activities supporting a categorization of OR faces and SR faces, respectively. Whole-brain analyses of BOLD responses to non-target OR faces revealed significant RS (Alt-Cond > Rep-Cond contrast) of activities in the PCC (Montreal Neurological Institute (MNI) coordinates x, y, z = 3, −67, 28) and medial prefrontal cortex (mPFC; −6, 53, 7, all activations were identified using a combined threshold of voxel level $P < 0.001$ uncorrected and cluster level $P < 0.05$ familywise error (FWE)-corrected; Supplementary Fig. 8). To further assess distinct patterns of RS of brain activity related to OR and SR faces, we conducted a whole-brain ANOVA to compare RS of neural responses related to OR and SR faces ((Alt-Cond versus Rep-Cond)$_{OR\,faces}$ versus (Alt-Cond versus Rep-Cond)$_{SR\,faces}$). The results showed significantly greater RS effects of PCC activity (3, −67, 25) to OR faces compared with SR faces.

Next, we conducted separate whole-brain analyses to examine RS of neural responses to faces of each racial group. This identified significant RS of neural responses to black faces in the PCC (−3, −70, 28) and the mPFC (3, 62, 1) (Fig. 5a), and to Asian faces in the left temporoparietal junction (LTPJ; −57, −46, 43) (Fig. 5c) only when comparing Asian faces alternately displayed with black faces in the Alt-Cond with those in the Rep-Cond (using a combined threshold of voxel level $P < 0.001$ uncorrected and cluster level $P < 0.05$ FWE-corrected). Unexpectedly, whole-brain analyses of BOLD responses to white faces failed to show a significant RS effect when using the same threshold. Because the ERP results in experiments 1–3 showed reliable RS effects on neural responses to white faces, we explored the RS of BOLD responses to white faces using a lenient threshold (voxel level $P < 0.005$ uncorrected, voxel number > 70). This analysis revealed RS of BOLD responses to white faces in the PCC (6, −64, 22) and left superior frontal cortex (LSFC; −12, 35, 55) (Fig. 5b).

A possible explanation for the weakened RS of BOLD responses to white faces in experiment 5a is that because each run during fMRI scanning included both Asian/black and Asian/white sessions, the presence of black faces might give prominence to similarity between white and Asian faces, and thus weakened the categorization of white faces as OR faces in Chinese participants. We tested this account in experiment 5b by examining whether BOLD responses related to the categorization of white faces would be more salient when Chinese participants perceived only white and Asian faces. We scanned an independent sample of Chinese participants ($N = 40$) using fMRI and stimuli and procedures similar to those for experiment 5a, but only showed Asian and white faces to participants (see Methods, Supplementary Table 7 for behavioural results). The whole-brain analyses of BOLD responses to white faces revealed significant RS effects on activities in the PCC (−12, −55, 16) and the LSFC (−33, 20, 49) (Fig. 5d) at the threshold of a voxel level $P < 0.001$ uncorrected and cluster level $P < 0.05$ FWE-corrected, thus replicating the results obtained in experiment 5a.

We also analysed effective connectivity between brain regions that showed significant RS effects for black faces and white faces using dynamic causal models (see Supplementary Methods 1 for details). The dynamic causal model results of black faces showed that repetition of black faces significantly modulated within-region connectivity in both the mPFC and the PCC and between-region connectivity from the mPFC to the PCC (see Extended Data Fig. 6; Supplementary Results 2), which suggested that the RS of PCC activity to black faces was possibly attributable to modulatory input from the mPFC. By contrast, repetition of white faces only significantly modulated within-region connectivity for the PCC. Together, these fMRI results indicate that distinct neural networks are involved in the categorization of OR and SR faces and of different subcategories of OR faces.

We further estimated whether the brain regions in which neural responses to OR faces (or SR faces) showed RS effects overlap with those in which there were significant differential neural responses to OR faces compared with SR faces. We conducted conventional analyses of contrasting faces of two racial groups in experiment 5a. The results showed that the brain regions showing significant differential neural responses to OR versus SR faces did not overlap with those in which the RS of activity to OR or SR faces took place (Supplementary Fig. 9). Unlike previous fMRI studies that usually calculated the contrast between OR versus SR faces (or vice versa), RS effects of brain activity in our experiment were defined by comparing BOLD responses to the same set of faces in the Alt-Cond versus the Rep-Cond and reflected how neural responses decreased to OR faces (or SR faces) due to habituation. These results indicate that fMRI data analyses focusing on differential neural responses to OR versus SR faces (or vice versa) do not by all means reveal neural structures supporting mental processes that are common for faces of different individuals to support racial categorization of these faces.

Experiment 6: dynamic neural models of racial categorization. Our ERP and fMRI results from the above experiments discovered distinct time courses and different brain regions involved in the categorization of OR and SR faces. These results, however, did not allow us to integrate time and spatial characteristics of the brain activities or to scrutinize dynamic interactions between different nodes of the neural networks involved in racial categorization during the first hundreds of milliseconds after face onset. In experiment 6, we sought to construct dynamic neural models of racial categorization of faces of each racial group by recording 306-channel, whole-head anatomically constrained MEG signals from an independent Chinese sample ($N = 26$). The stimuli and procedures were the same to those used in experiment 5a, except that shorter stimulus durations (200 ms) and interstimulus intervals (250–550 ms) were used and there were 128 trials in each condition (see Methods, see Extended Data Fig. 7 for behavioural results).

Analyses of sensor-space and source-space MEG signals first validated the distinct time courses of racial categorization of OR and SR faces shown in our ERP results and greater RS of neural responses to OR faces than SR faces shown in our fMRI results (Extended Data Figs. 8 and 9; Supplementary Results 3). We then conducted separate cluster-based permutation t-tests (using a predefined threshold of $P < 0.025$, two-tailed, 10,000 iterations, and a cluster-level threshold of $P < 0.05$ with corrections for multiple comparisons) to examine the neural dynamics of categorization of faces of each racial group. The results showed significant RS of sensor-space signals to black faces at bilateral temporal sensors (140–200 ms; left cluster: $P = 0.042$, right cluster: $P = 0.043$, corrected; only cluster-level-corrected P values after further false discovery rate (FDR) correction across tests are subsequently reported) and at bilateral occipital, parietal and temporal sensors (230–400 ms; left cluster: $P = 0.0014$, right cluster: $P = 0.0014$). Significant RS of sensor-space signals were also identified at bilateral temporal and occipital and left parietal sensors for white faces (140–200 ms; left cluster: $P = 0.043$, right cluster: $P = 0.040$) and at the left temporal and parietal sensors for Asian faces (350–390 ms; cluster $P = 0.048$) (Fig. 6). These results are consistent with our ERP results in showing evidence for earlier racial categorization of OR than SR faces.

Next, we examined dynamic activity of the neural network involved in the categorization of Asian, black and white faces by conducting whole-brain source-level analyses using a predefined threshold of $P < 0.025$, one-tailed, 10,000 iterations, and a cluster-level threshold of $P < 0.05$ with corrections for multiple comparisons (one-tailed tests were used here to examine decreased neural responses to faces in the Rep-Cond versus the Alt-Cond). The results yielded significant clusters showing RS effects related to
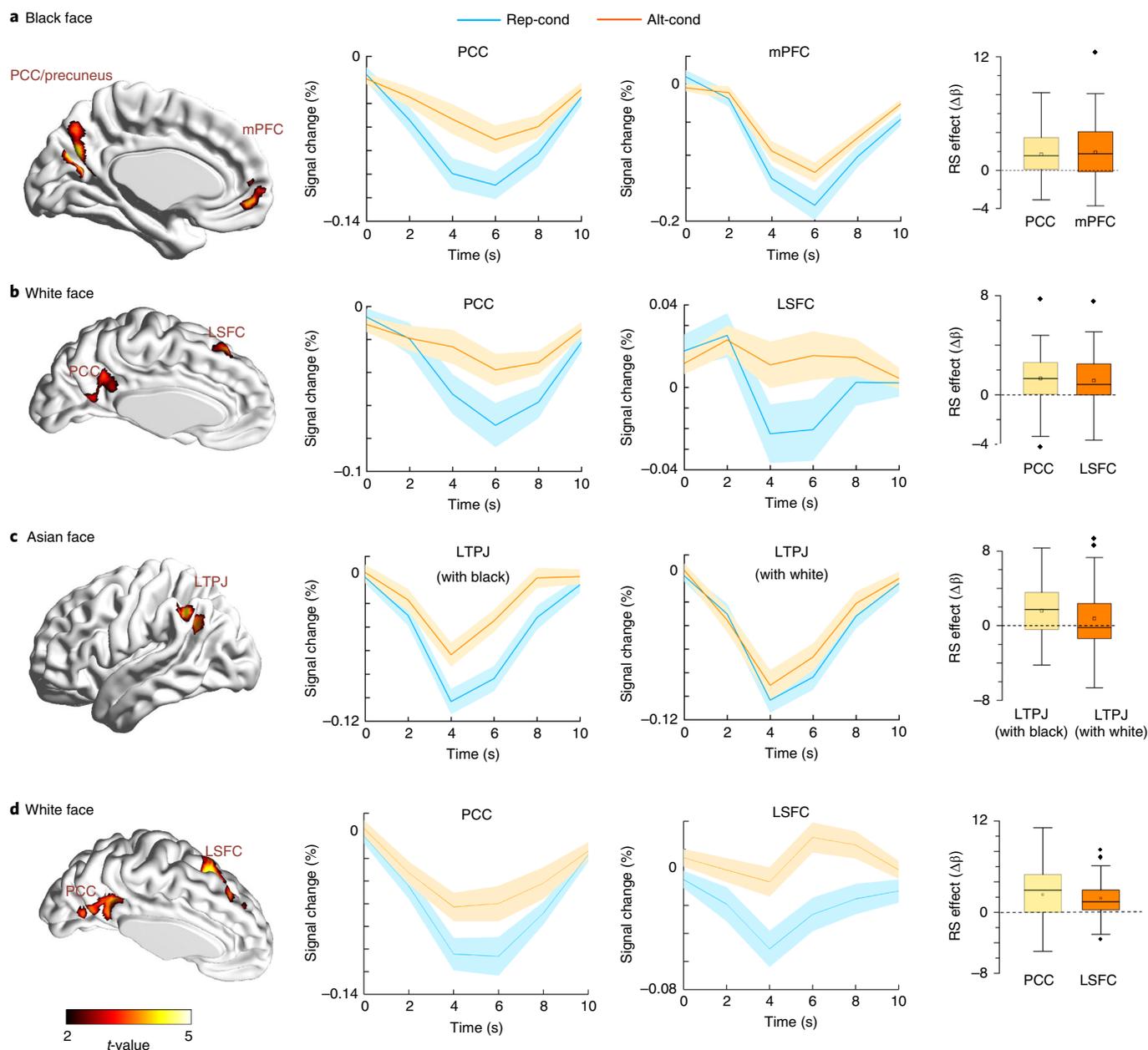
Fig. 5 | fMRI results of Chinese participants in experiments 5a and 5b. a–c, Results of experiment 5a ($N = 55$; 2 participants were excluded from analyses due to excessive head movement). **d**, Results of experiment 5b ($N = 40$). For all panels, the left column shows brain regions in which significant RS of BOLD responses occurred. The middle column illustrates time courses ($\pm$ s.e.m. shown as shading around the mean trace) of BOLD responses in these brain regions in the Alt-Cond and the Rep-Cond. The box plots in the right column illustrate the mean RS effects indexed by differences in β-values (Alt-Cond minus Rep-Cond) in each brain region. The plots show the quartiles (boxes), means (square inside boxes), medians (horizontal lines inside boxes), maximum and minimum excluding outliers (whiskers), and outliers (diamonds). RS effects on BOLD responses in **a**, **c** and **d** were identified and visualized using a combined voxel level threshold of $P < 0.001$ uncorrected and a cluster level of $P < 0.05$ FWE-corrected. RS effects on BOLD responses in **b** were explored and visualized using a threshold of $P < 0.005$ uncorrected and a voxel number > 70.

black faces at 140–200 ms in the right fusiform gyrus (RFG) (peak MNI coordinates: x, y, $z = 42$, -32, 19; cluster $P = 0.043$ corrected; only cluster-level-corrected P values are subsequently reported) and the left fusiform gyrus (LFG; -47, -58, -23) extending to the left anterior temporal lobe (LATL (-53, -2, -42); cluster $P = 0.003$) (Fig. 6). RS effects of neural responses to black faces were also identified in the mPFC (-3, 58, 15; cluster $P = 0.017$) at 230–400 ms. The results also revealed a marginally significant cluster showing a RS effect at 350–390 ms related to Asian faces in Asian/black sessions in the LTPJ (-38, -57, 54; cluster $P = 0.067$).

Given that our fMRI results and sensor-space MEG results had identified reliable RS of neural responses to white and Asian faces in specific brain regions and time windows, we further conducted region-of-interest (ROI) analyses to examine RS effects on source-space MEG signals. ROIs were defined based on our fMRI results and anatomically based regions (see Methods). Cluster-based permutation tests were performed within each mask (using a predefined threshold of $P < 0.025$, one-tailed, 10,000 iterations, and a cluster-level threshold of $P < 0.05$ after further FDR correction across tests). The results revealed significant RS of activity in responses to white
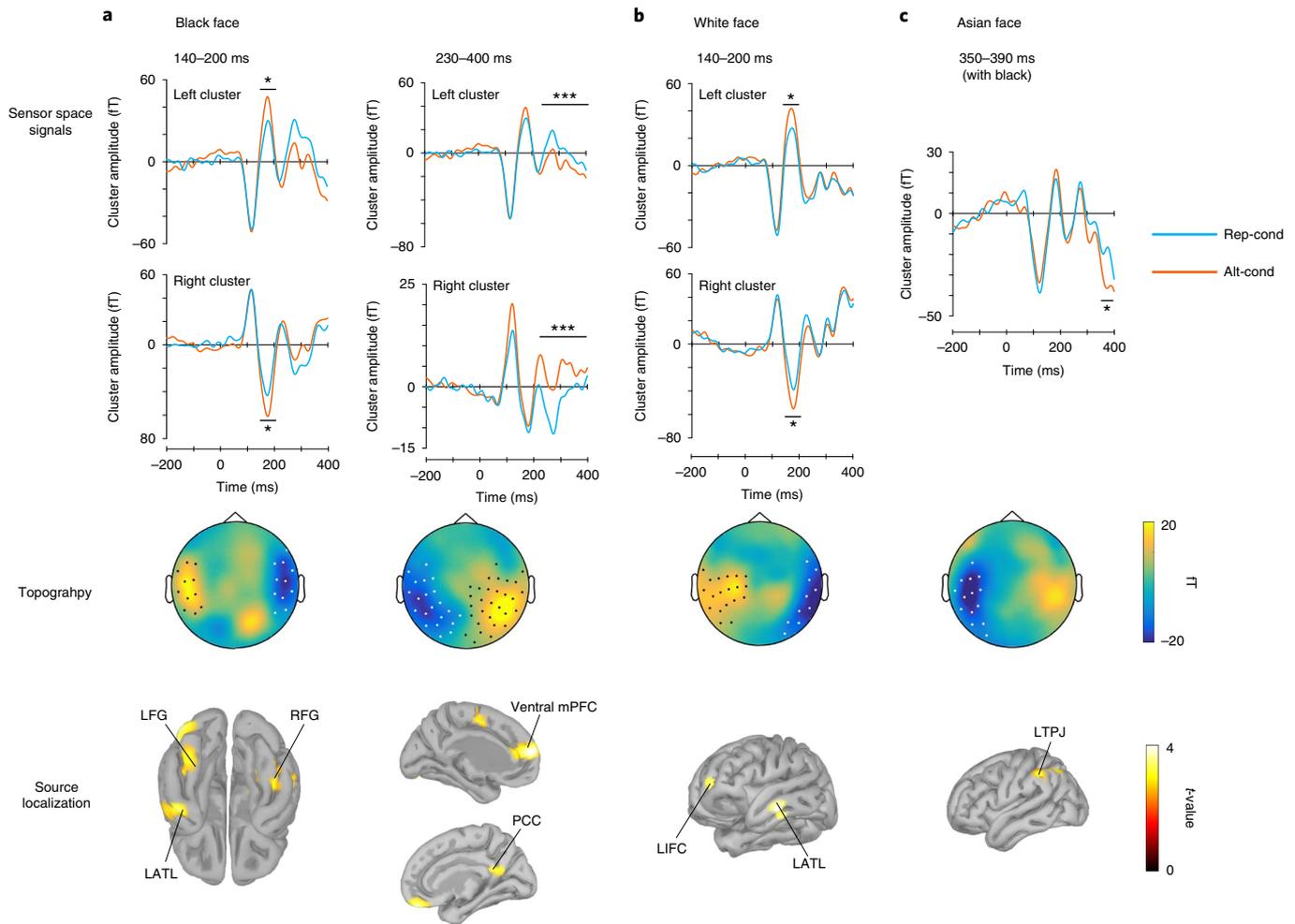
Fig. 6 | MEG results of Chinese participants in experiment 6. a–c, Results relate to black (**a**), white (**b**) and Asian faces (**c**). The upper panels illustrate the grand average sensor-space MEG signals averaged across the sensors in significant clusters (*N* = 26). The middle panels show topographies of the RS effects (Alt-Cond minus Rep-Cond) in each time window (dots represent sensors contributing to the significant clusters). The lower panels show source brain regions that contribute to the RS effects on sensor-space signals in each time window. The RS of MEG signals to all faces were identified using a threshold cluster of *P* < 0.025 (two-tailed for sensor-space and one-tailed for source-space) and corrected for multiple comparisons. A lenient threshold of 30 contiguous vertices, *P* values < 0.005 uncorrected, were used only for source-space visualization.

faces in the left inferior frontal cortex (LIFC (−24, 59, −4); cluster *P* = 0.045) and the LATL (−65, −4, −28; cluster *P* = 0.040) at 140–200 ms and in the PCC (1, −30, 45; cluster *P* = 0.015) at 300–400 ms. Significant RS effects were also observed in PCC/precuneus activity (2, −58, 12; cluster *P* = 0.040) at 230–400 ms for black faces and in the LTPJ (−38, −57, 54; cluster *P* = 0.015) at 350–390 ms for Asian faces in Asian/black sessions.

To assess the fine spatiotemporal patterns of RS of source-space signals, we extracted the time courses of source-space signals from the regions showing significant RS effects. Similar to the protocol presented above, one-tailed cluster-based permutation *t*-tests (using a predefined threshold of *P* < 0.025, one-tailed, 10,000 iterations) were conducted to verify decreased neural responses in the Rep-Cond compared with the Alt-Cond (all cluster *P* values < 0.05 after further FDR correction across regions). The results revealed that for black faces, RS of neural responses (that is, decreased activity in the Rep-Cond versus the Alt-Cond) emerged at 150 ms after face onset in the LFG, RFG and LATL, but later in the mPFC (around 200 ms) and the PCC (around 300 ms). For white faces, significant RS of LATL and LIFC activity was detected around 150 ms, followed by RS of PCC activity around 230 ms. For Asian faces (paired with black faces), significant RS of LTPJ activity was detected around

340 ms (Fig. 7; see Supplementary Table 8 for statistical details). These results further dissociate the different brain regions involved in the categorization of OR and SR faces with distinct time courses.

To further examine the dynamic interactions among the brain regions involved in the racial categorization of faces, we performed a Granger causality analysis (GCA) of RS effects (Alt-Cond minus Rep-Cond) on averaged source-space signals (0–400 ms) in response to faces of each racial group to probe the directional information flow between the source pairs identified above (see Methods). The temporal profiles of source signals in response to black faces suggest a two-stage model given the early RS effect exhibited in the LFG, RFG and LATL (before 200 ms) and the late RS effect in the mPFC and PCC (after 200 ms). Indeed, the GCA results showed significant mutually interactive connectivity among the regions that exhibited the early RS effect (that is, the RFG, LFG and LATL) and significant input from the LATL, LFG and RFG to the mPFC and PCC. The connectivity between the mPFC and the PCC was unidirectional, and the PCC further exhibited input to the LFG, which might serve as a top-down feedback pathway (*P* values < 0.05, Bonferroni-corrected; Fig. 7a). These results, together with the temporal profiles of source signals, suggest that there is an early feed-forward information flow from the LFG–RFG–LATL to the
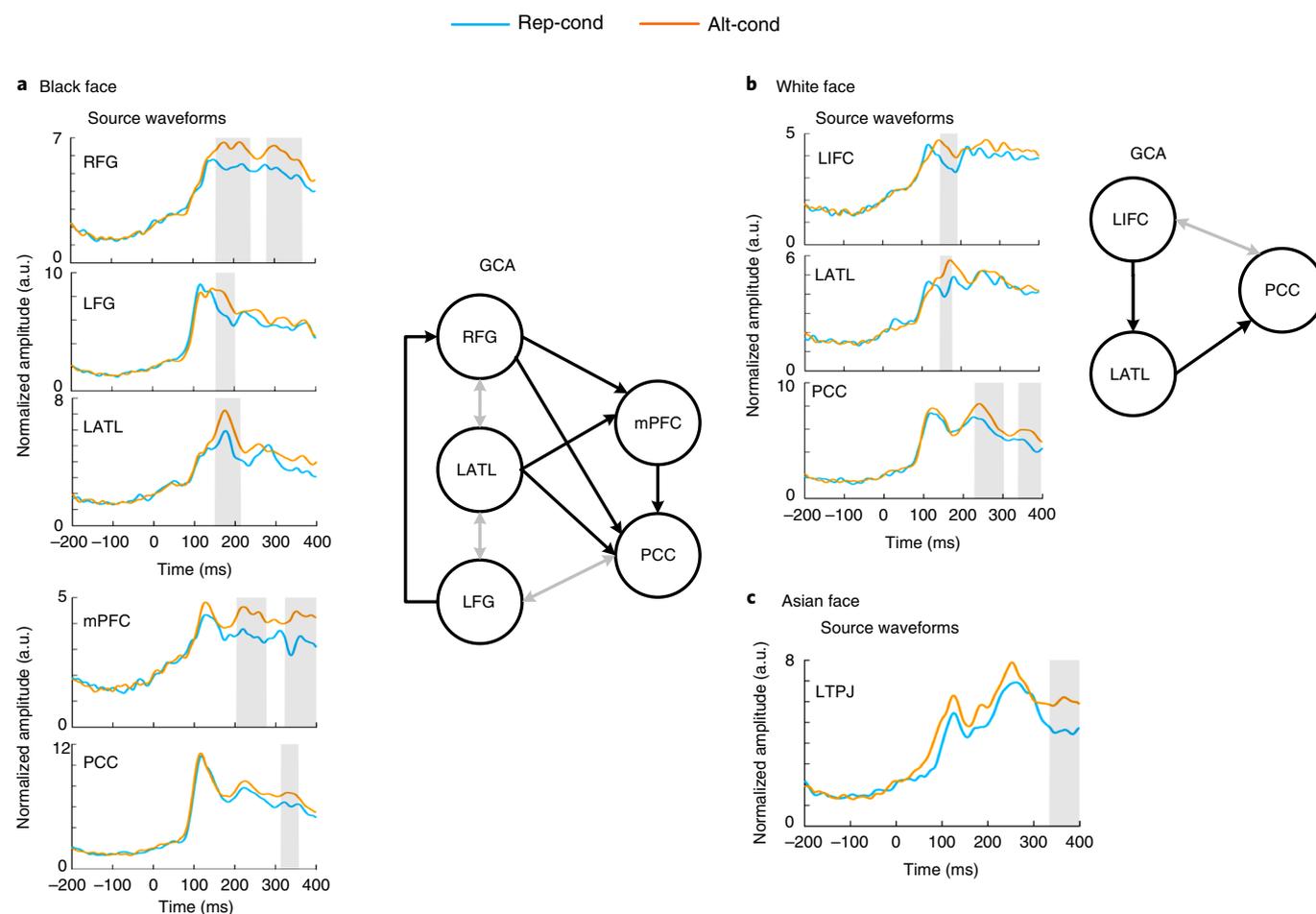
Fig. 7 | Dynamic neural models of racial categorization revealed in experiment 6. a–c, RS effects on source-space MEG signals and dynamic models of categorization processes for black (**a**), white (**b**) and Asian faces (**c**). The grand average of source-space MEG signals in the brain regions included in the GCA are shown on the left of each panel, and dynamic models of categorization processes are shown on the right of each panel (note, there is no model for Asian faces). Time windows showing reliable RS effects on source-space MEG signals are marked with grey bars ($N = 26$, cluster P values < 0.05, FDR-corrected). Black lines and arrows in the dynamic models indicate unidirectional causal influences. Grey lines and arrows indicate bidirectional causal influences. Significance was assessed using a threshold of $P < 0.05$, Bonferroni-corrected for GCA. a.u., arbitrary unit.

mPFC–PCC and late feedback input from the mPFC–PCC to the fusiform during the categorization of black faces.

The temporal patterns of source signals for white faces also suggest a two-stage model given the early RS effect in the LATL and LIFC (before 200 ms) and the late RS effect in the PCC (after 200 ms). GCA of source-space signals in response to white faces suggested feed-forward connectivity from the LATL and LIFC to the PCC and feedback connectivity from the PCC to the LIFC (Fig. 7b). The results also suggested that there is an early information flow from the LIFC to the LATL. These results provided further evidence that the categorization of OR faces engages two-stage dynamic activities in neural networks consisting of multiple interactive brain regions, whereby two different networks are engaged in spontaneous categorization of two subgroups of OR faces (that is, black and white faces). By contrast, categorization of SR faces recruits a different and simple network at a later time window.

To estimate whether the neural dynamics of racial categorization varied along implicit negative attitudes of individuals towards OR faces, we asked the participants in experiment 6 to complete the implicit association test (IAT)[52] in two sessions (that is, Asian/black sessions and Asian/white sessions) after MEG recording. The D score, which is calculated on the basis of the response latencies of the participants[53], reflected implicit negative attitudes towards OR faces. One-sample t-test revealed that the D score was significantly larger than 0 in Asian/black sessions (0.453 ± 0.44, $t(25) = 5.310$, $P < 0.001$, 95% CI = 0.28, 0.63, Cohen's $d = 1.03$), indicating that compared with Asian faces, black faces were associated with negative attitudes. However, the D score did not differ significantly from 0 in Asian/white sessions (0.00 ± 0.44, $t(25) = -0.009$, $P > 0.9$, 95% CI = −0.18, 0.18, Cohen's $d = 0.002$), thus showing that there is no evidence for implicit negative attitudes towards white faces in our participants. We calculated correlations between D scores and RS effects on both sensor-space and source-space signals, but did not find significant results (see Supplementary Table 9 for statistical details), thus showing that there is no evidence for associations between the neural dynamics of racial categorization and implicit negative attitudes towards OR faces.

Neural processes of racial categorization and recognition deficit of OR faces. Behavioural research has shown greater false alarm rates of recognizing OR faces (versus SR faces)[5] and that such an impaired OR face recognition (that is, the 'other race effect') has been attributed to both an increased individuation of SR faces and an enhanced categorization of OR faces[18,19]. Based on these models and our findings of a P2 RS effect related to the categorization of OR faces in experiments 1–3, one may predict that individuals with
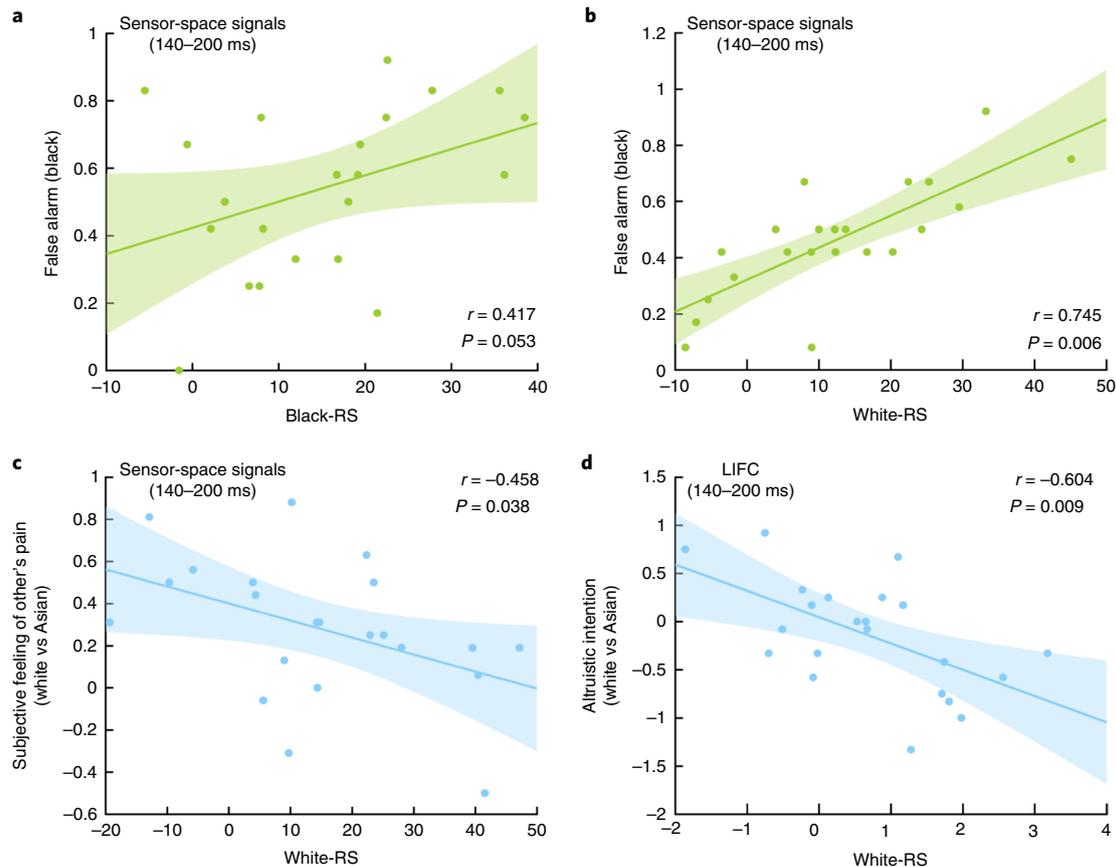
Fig. 8 | Coupling between neural categorization of OR faces and racial biases in cognition and altruistic intension in experiment 6. a,b, Correlations between early RS effects on sensor-space MEG signals to black (**a**) and white (**b**) faces and false alarm rates in the face recognition task. Greater RS of sensor-space MEG signals to white faces at 140–200 ms predicted more false alarms when recognizing old/new white faces ($N=22$, $r=0.745$, $P=0.006$, 95% CI = 0.42, 0.92, FDR-corrected). A similar trend was observed for black faces ($r=0.417$, $P=0.053$, 95% CI = −0.05, 0.78, FDR-corrected). **c**, The correlation between early RS effects on sensor-space MEG signals to white faces and subjective feelings of pain of white (versus Asian) faces. Stronger RS of sensor-space MEG signals to white faces at 140–200 ms predicted weaker subjective feelings of pain of white (versus Asian) faces ($N=22$, $r=-0.458$, $P=0.038$, 95% CI = −0.77, −0.04, FDR-corrected). **d**, The correlation between RS of source-space LIFC responses to white faces and altruistic intentions to help white (versus Asian) individuals who suffer from painful stimulations. Stronger RS to white faces in LIFC predicted less intention to help white (versus Asian) models ($N=22$, $r=-0.604$, $P=0.009$, 95% CI = 0.22, 0.81, FDR-corrected).

greater RS of neural responses to OR faces (that is, an enhanced categorization of OR faces) should exhibit worse recognition of OR faces. To test this prediction, after MEG recording, we asked the participants in experiment 6 to complete a face recognition task to discriminate between a previously shown subset of faces and a subset of novel faces (see Methods). Similar to previous published findings[5], the recognition performance showed higher false alarm rates for black/white faces compared with Asian faces (see Supplementary Results 3 for statistical details). Moreover, nonparametric Spearman's rank correlation analyses revealed that greater RS of sensor-space MEG signals to white faces at 140–200 ms predicted more false alarms when recognizing old/new white faces ($r=0.745$, $P=0.006$, 95% CI = 0.42, 0.92, FDR-corrected; Fig. 8a). A similar trend was observed for black faces ($r=0.417$, $P=0.053$, 95% CI = −0.05, 0.78, FDR-corrected; Fig. 8b). The analyses of source-space MEG signals also showed that greater RS of neural responses to black faces in the LFG and LATL predicted more false alarms during recognition of black faces (see Extended Data Fig. 10, Supplementary Results 3 for statistical details). These correlation results are consistent with the psychological models[18,19] and provide neuroimaging evidence for coupling between neural processes underlying the racial categorization of OR faces and impaired recognition of OR faces.

Neural processes of racial categorization and altruistic intention. Finally, we tested the predictions of a psychological model whereby an enhanced categorization of OR faces results in decreased empathy and social decision-making related to the suffering of OR individuals[10]. The participants in experiment 6 were shown Asian and white faces with painful or neutral expressions (these faces were from a previous study[14]) and asked them to report their subjective feelings of pain of these faces (measured using a Likert-type scale). Participants were also presented with video clips of Asian and white models receiving painful or neutral stimulations (used in previous work[54]) and had to report to what degree they intended to help these models (see Methods). Difference rating scores (white faces (pain − neutral) minus Asian faces (pain − neutral)) were calculated to measure racial biases in empathy and altruistic intention. Nonparametric Spearman's rank correlation analyses showed that stronger RS of sensor-space MEG signals to white faces at 140–200 ms predicted weaker subjective feelings of pain of white (versus Asian) faces ($r=-0.458$, $P=0.038$, 95% CI = −0.77, −0.04, FDR-corrected; Fig. 8c). Moreover, stronger RS of source-space MEG signals to white faces in the LIFC predicted less intention to help white (versus Asian) models ($r=-0.604$, $P=0.009$, 95% CI = 0.22, 0.81, FDR-corrected; Fig. 8d). These results suggest possible neural mechanisms that associate racial categorization with racial bias in

psychological underpinnings (for example, empathy and altruistic motivation) of prosocial behaviour[10].

Discussion

Other racial groups constitute part of the natural environment[55] in which individuals learn to categorize people into different racial groups during development[56]. Here, across six experiments and using multimodal neuroimaging techniques, we investigated the neural processes supporting spontaneous racial categorization by examining the adaptation of neural responses to faces in a task requiring attention to individual identity. Our methods for estimating neural adaptation were different from previous imaging studies that directly compared neural responses to OR versus SR faces[23,36]. Using EEG, fMRI and MEG, we examined RS effects of neural responses by comparing brain activities in response to faces of one racial group (that is, Asian, black or white) in an Alt-Cond and a Rep-Cond. This approach allowed us to disentangle time courses of neural activities in specific neural networks involved in the spontaneous categorization of faces of each racial group. Our findings provide a neurobiologically grounded framework for constructing distinct psychological models of spontaneous categorization of faces from different racial groups and for understanding racial bias in social behaviour.

Our pilot experiment provided behavioural evidence that Asian, black and white faces used in the subsequent imaging experiments were perceived as three separate categories and that black faces were more densely clustered than white/Asian faces. These results replicate previous findings using similar measures[57] and implicate distinct neural underpinnings of categorization of faces of different racial groups. Consistent with this proposition, our EEG results in experiment 1a revealed RS of early neural activity (that is, the frontal P2 at 140–200 ms) in response to OR faces. RS of neural responses to SR faces, however, emerged later in the N2 time window (200–400 ms) when SR faces were presented alternately with black faces. The distinct temporal patterns of adaptation of neural responses to OR and SR faces were replicated using different sets of face stimuli (experiment 1b), after controlling for influences of gender categorization of faces (experiment 2), and in both Chinese (experiments 1 and 2) and white (experiment 3) participants. These results provide empirical evidence for the psychological models of race perception[10,18,19] that assume enhanced or early categorization of OR compared with SR faces by assessing the exact timing of neural processes of OR and SR faces. Experiment 4 showed that face inversion weakened both P2 and N2 RS effects. This finding implies that perceived skin tone is not sufficient to activate the neural processes underlying racial categorization, whereas facial configural information seems to be necessary for the neural processes that support both early OR categorization and late SR categorization.

The approach we used in assessing the RS of neural responses to faces of each racial group is different from the methodology that compared neural responses to faces of two racial categories and revealed larger N1 and P2 amplitudes to black than white faces in white participants[20–25] (the same method of data analysis resulted in similar outcomes in our work). The early RS effect demonstrated in our experiments occurred in the P2 time window rather than in the N1 time window. Because the N1 effect has been thought to reflect early attentional orientation to novel targets such as black faces[23], our results suggest that early attentional orientation to each individual black face might take place regardless of how often these faces are perceived and grouped into one racial group. Moreover, the P2 RS effect demonstrated across our EEG experiments was specific to OR faces, whereas the RS of N2 amplitudes was specific to SR faces. Thus, the approach we used in investigating the RS of neural responses to faces disentangled separate time courses of the spontaneous categorization of OR and SR faces.

Our fMRI results in experiments 5a and 5b unravelled that for Chinese participants, RS of neural activities occurred in the mPFC and PCC for black faces, but in the PCC and LSFC for white faces. The analysis of effective connectivity between these brain regions further suggested that repetition of black faces modulated the information flow from the mPFC to the PCC. Unlike the RS of neural responses to black and white faces, reliable RS effects of neural activity in response to Asian faces was manifested only in the LTPJ. These fMRI results reinforced our ERP findings by unveiling distinct neural networks that support the spontaneous racial categorization of OR and SR faces and of different subcategories of OR faces. Interestingly, some brain regions in which responses differed significantly when contrasting black versus white faces in previous research (for example, the amygdala and the anterior cingulate)[26,27,29] did not show significant RS effects on neural responses to either black or white faces in our work. Greater amygdala and anterior cingulate activities to OR faces versus SR faces are assumed to reflect encoding of emotional salience of OR (for example, black) faces and monitoring of conflict between prepotent feelings and conscious intentions to respond fairly[36]. The fact that amygdala and anterior cingulate activities failed to show significant RS effects in our work implicates that encoding of emotional salience and conflict monitoring related to each individual black face might occur similarly in the Alt-Cond and the Rep-Cond and therefore be independent of spontaneous racial categorization. In our work, different patterns of brain activations relating to OR versus SR faces and different RS effects relating to neural responses to faces of each racial group were obtained when the same (that is, one-back) task was used in both the Alt-Cond and the Rep-Cond and for both SR and OR faces. It appears that our experimental design and data analyses, rather than the task, were pivotal for assessing neural responses related to racial categorization of faces.

Our MEG results obtained in experiment 6 validated the EEG results in experiments 1–3 by showing early RS effects of neural responses to OR faces and late RS effects of neural responses to SR faces. More importantly, the analysis of source-space MEG signals further revealed neural networks in which the activity showed RS effects pertaining to black, white or Asian faces within 400 ms after face onset and suggested distinct time courses and dynamic interactions of neural activities in these networks. Specifically, categorization of black faces—one subcategory of OR faces for Chinese and white participants—may be inferred form the spatiotemporal pattern of activities characterized by early activations of the fusiform and LATL and feedforward input from these regions to the mPFC and PCC, which are followed by activations of the mPFC and PCC and backward influences of the PCC on fusiform activity. The two-stage dynamic neural model of categorization of black faces bolsters the emerging cognitive functions of the related brain regions and reinforces the dynamic and interactive models of social perception[58,59]. Our MEG results suggest that early processes of categorization of black faces may include coding of common facial features of black faces in the fusiform gyrus[60] and activation of the concept 'black' as a social category in the LATL[61]. Configural and category information obtained at the early stage may further be exported to the mPFC to facilitate the processing of person knowledge and stereotype[62,63] by collecting social category information related to blacks represented in the LATL[16,63], which might further guide the retrieval of relevant information from episodic memory in the PCC/precuneus[64].

Our MEG results suggest a different dynamic neural model for the categorization of white faces—another subcategory of OR faces for Chinese participants. This model suggested early employment of the LATL and LIFC and later involvement of the PCC. The neural dynamics of categorization of white faces may start from early activation of a racial category concept (for example, white) in the LATL and LIFC, which are also activated during social judgments

of SR and OR faces[65]. This might be followed by memory retrieval related to white faces in the PCC/precuneus[64] and its feedback link with the LATL. These results suggest that although both black and white faces are perceived as OR faces for Chinese participants, racial categorization of black and white faces may depend on different dynamic neural processes in distinct neural networks.

A simple dynamic neural model characterizes the processes that are common for faces of different individuals of one's own racial group in a late (for example, N2) time window and are associated with the LTPJ. The LTPJ is usually activated when perceiving goal-directed movement, inferring others' beliefs[66] and during processing of social-category-related behavioural information[67]. Our findings highlight a new functional role of the LTPJ in the spontaneous processing of SR faces that might emphasize mental state inference and can be enhanced by the presence of black faces, possibly due to the negative attitude and stereotype threat associated with blacks[43–45,68]. Overall, the dynamic models of categorization of Asian, black and white faces emerging from our multimodal brain imaging results go beyond previous research that focused on differential neural processes of OR and SR race faces using either EEG or fMRI measures[23,36] by uncovering categorization processes of faces from each racial group. Our results indicate that faces from different racial groups are categorized using separate neural classifiers and provide a neural basis for the psychological models of racial bias in recognition and emotion[10,18,19].

Both Chinese and white participants in the current work were tested in the same social environment where black and white faces are encountered less frequently than Asian faces. Nevertheless, the Chinese and white participants showed mirror patterns of RS effects on the P2 and N2 amplitudes in responses to Asian and white faces. These results demonstrate that racial relationships between observers and perceived faces rather than the frequency of exposure to faces of different races in the testing environment determine the processes of categorization of faces of a specific racial group. In addition, despite differences in life experiences and growth environments between the Chinese and white participants recruited for our work, the two subject groups showed the same pattern of neural responses related to an early categorization of OR faces (that is, P2 RS effect), but a late categorization of SR faces (that is, N2 RS effect). Thus, distinct temporal procedures of categorization of OR and SR faces are common for individuals from different societies and support the psychological models that emphasize early categorization of OR faces[10,18,19].

Our MEG results revealed that the neural activities underlying racial categorization at 140–200 ms after face onset predicted a recognition deficit of OR faces. These results support the view that early racial categorization provides the precondition for racial bias in memory[18,19] and offer a potential neural mechanistic account. Earlier work examined the association between differential neural responses to OR versus SR faces and differential memory performance of OR versus SR faces[28,69] and interpreted impaired OR face recognition based on enhanced configural encoding[70] and more in-depth processing of SR faces[25]. Recent research, however, failed to show evidence for a correlation between behavioural measures of outgroup deindividuation and differential adaptation of fusiform responses to unique identities of OR versus SR faces[71]. Unlike the previous approach, our results related early processing of OR faces as a group within 200 ms after face onset to the impaired recognition of OR faces as individuals. The source-space MEG signals further suggest coupling between memory deficits for black faces and common neural processes of black faces as a stimulus class in both the LFG and the LATL. These two brain regions are functionally connected during face processing to code information about facial identity[72] and constitute a key network involved in the integration of facial cues with social-conceptual information to form initial perceptions of others[59]. Our results suggest that categorization of different

black faces as a social group by integrating common facial and social-conceptual information may contribute to deficits of processing individual identity of these faces. However, this neural mechanistic account may be specific to the recognition of black faces, because source-space MEG signals failed to predict false alarm rates related to white faces. Future work should clarify whether recognition deficits pertaining to black and white faces have different neural underpinnings, although both black and white faces were categorized as OR faces for Chinese participants. Our sensor-space MEG signals also suggest associations between early racial categorization and decreased empathy for pain for OR individuals and relevant intentions for prosocial decisions. These results reinforce the psychological model that assumes a link between enhanced categorization of OR faces and reduced empathy for their suffering and decreased altruistic intention[10]. Together, our findings suggest that early spontaneous categorization of OR faces plays a key role in generating racial bias in cognition, emotion and social decision-making, and should be taken into consideration when developing programmes for intervention of racial bias in behaviour.

Previous behavioural findings of implicit negative attitudes towards OR individuals[6,7] led to the question of whether the distinct spatiotemporal patterns of neural adaption relating to faces of different racial groups reflect the impact of implicit negative attitudes or stereotype. Indeed, some of the brain regions identified in our work (for example, the mPFC) have been shown to activate during tasks that require inference of personal traits or stereotype-based judgments of black (versus white) people[63,73]. Our behavioural measures also showed evidence for implicit negative attitude towards black but not white faces in participants in experiment 6. However, distinct neural dynamics of racial categorization of Asian, black and white faces revealed in the current work are unlikely to merely reflect neural coding of implicit negative attitudes. Distinct spatiotemporal patterns of neural adaption pertaining to white and Asian faces were evident in our Chinese participants who did not show implicit negative attitudes towards white faces. Future research may use the paradigm developed here to further address other cognitive and affective processes that distinguish Asian, black and white faces during racial categorization.

In conclusion, our findings demonstrate that the brain engages distinct neural dynamics to sort faces into different racial categories. While recent research probed possible intertwined neural structuring of different social categories (for example, race, gender and emotion)[59,74], our work teased apart separate neural processes underlying racial categorization of subgroups of faces. Our findings implicate that distinct neural mechanisms may evolve in the human brain, even for the categorization of faces of subgroups, along one social dimension (for example, race). Our brain imaging results break through the SR/OR dichotomy of racial categorization by unravelling distinct dynamic neural processes involved in the categorization of faces into multiple racial groups. The distinct neural models of racial categorization discovered in our work present a new scientific perspective on racial bias in social behaviour by uncovering the links between the neural underpinnings of racial categorization and racial biases in cognition, emotion and decision-making. Our work assessed potential associations between racial categorization and racial bias in the psychological underpinnings of prosocial behaviour, but limited to white faces. Future research should further test such associations for faces of other racial groups. Finally, it should be noted that the current work tested two ethnic samples (that is, Chinese and white participants), and RS of neural responses to SR faces was observed in a specific condition (that is, when being presented alternately with black faces). Whether the current findings can be generalized to other ethnic samples and faces of other races remains unclear and should be tested in future research. By integrating the RS paradigm with multimodal brain imaging, our work opens a new avenue for social neuroscience research to scrutinize

the complicated mechanisms of social cognition by examining neural dynamics with both good timing and spatial resolution.

Methods

Participants. This study included the following participants: 57 Chinese students for the pilot experiment (28 males, mean age ± s.d. = 21.5 ± 2.3 years); 38 Chinese students for experiment 1a (17 males, 20.5 ± 1.1 years); 38 Chinese students for experiment 1b (19 males, 21.7 ± 2.7 years); 36 Chinese students (18 males, mean age ± s.d. = 21.4 ± 2.2 years) for Asian/black sessions in experiment 2; 31 Chinese students (15 males, mean age ± s.d. = 21.5 ± 2.1 years) for Asian/white sessions in experiment 2; 35 white students in experiment 3 (19 males, 22.0 ± 2.1 years, 3 participants did not complete white/black EEG sessions); 35 Chinese students for experiment 4 (17 males, 19.5 ± 1.1 years); 55 Chinese students in experiment 5a (29 males, 21.6 ± 1.9 years, 2 participants were excluded from data analyses due to head movement during scanning); 40 Chinese students in experiment 5b (20 males, 22.5 ± 2.1 years); and 26 Chinese students in experiment 6 (14 males, 22.7 ± 2.2 years). All Chinese students were born and raised in China. All white students were self-identified as white and had lived in China for 2 weeks to 1 year when being tested. Among the white participants, there were 7 from the United States, 2 from Australia and 26 from Europe. All participants were tested in Beijing, China. Experiments 5a and 5b had larger sample sizes relative to the EEG and MEG experiments because the long interstimulus intervals used might weaken the RS effect on BOLD signals. All participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric diagnoses. This study was approved by the local Research Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University. All participants provided written informed consent after the experimental procedure had been fully explained. Participants were reminded of their right to withdraw at any time during the study and were paid for their participation.

Stimuli. Stimuli consisted of three sets (Asian, black and white) of faces with neutral expressions. Sixteen Asian (8 males) and 16 white (8 males) faces were used from our previous study[13], 16 black faces (8 males) were used from the MR2 database[75]. These stimuli were used in the pilot experiment, experiments 1a and 3–6 (Supplementary Fig. 1). Another three sets (Asian, black and white) of faces, cropped to remove hair, were selected from the MR2 database[75] and the Chicago face database (v.2.0.3)[76] for experiments 1b and 2. All faces used in our work showed neutral expressions. The luminance levels of Asian, black and white faces were adjusted and matched. Upright faces were used in all experiments, except for experiment 4, which used inverted faces. Each face subtended a visual angle of 3.8° × 4.7° at a view distance of 60 cm in the pilot experiment and in all EEG experiments. Each face subtended a visual angle of 8.60° × 13.61° at a view distance of 80 cm in the fMRI experiment. Each face subtended a visual angle of 6.9 × 9.9° at a view distance of 75 cm in the MEG experiment.

Procedures of behavioural tasks. *Similarity rating task.* The multidimensional space framework posits that each face is represented in a multidimensional feature space (face space), and the concept of cluster density is used to describe how many neighbours a face has in the face space[77,78]. Experiences with SR faces result in distributed representations of SR faces in the face space defined by facial feature dimensions, which make it easy to discriminate individual SR faces. By contrast, OR faces are densely clustered in the face space and leads to difficulty in discriminating individual OR faces. In the pilot experiment, we adopted a behavioural paradigm from previous work[57] to examine how SR and OR faces are represented in the face space and, in particular, to assess whether SR and OR faces are perceived into different clusters. The face stimuli used in the EEG, fMRI and MEG experiments were employed in the similarity rating task. During the similarity rating test, each participant was asked to rate all the 300 unique same gender face-pairs presented in a random order on a 9-point scale (where 1 represents not similar at all and 9 represents extremely similar) with the following instructions, which were similar to those used in previous research[57]: "You will be shown pictures of pairs of faces. There are Asian, white and black faces in the experiment. Facial characteristics may differ across racial groups but some faces may still be quite similar even they are from different racial groups. Faces of the same race may also differ markedly in terms of similarity. Try to consider each pair of faces as two individuals, regardless of race, and answer the following question. How similar are the two faces?". The left/right positions of each face within face pairs were counterbalanced across participants. Each face pair was presented until a participant finished a similarity rating by giving a rating score. We calculated the mean distance between each face and other SR faces and the mean distance between each face and OR faces in a four-dimensional face space.

Old/new face recognition task. An old/new face recognition task[57] was conducted to estimate the performance of participants to recognize SR and OR faces after MEG recording in experiment 6. The face stimuli used in the memory task consisted of three sets of faces, including 24 Asian faces, 24 black faces and 24 white faces, cropped to remove hair, which were distinct from the faces used in the MEG recording sessions. These faces were taken from the MR2 database[75] and

the Chicago face database (v.2.0.3)[76]. Faces of each race category were randomly divided into two subsets (half male and half female faces) for each participant. During the encoding stage of the recognition task, a subset of faces from each racial category was presented in a random order. Participants were asked to passively view and to remember the faces. Each face was presented for 1 s followed by 400–600 ms of fixation. After the encoding stage, participants were asked to complete a 5-min calculation task as distraction. After the calculation task, participants were presented with all learned faces and an equal number of new faces from the three racial categories in a random order. Each face was displayed on the screen until participants indicated whether they had previously seen this face or not by pressing one of two buttons. Because previous a meta-analysis[5] has shown that memory deficits for OR faces compared with SR faces are most prominent in false alarm rates, we calculated false alarm rates for each set of faces to examine the difficulty in recognizing individual faces of a specific racial category.

Empathy rating task. To assess subjective feelings of pain shown in OR and SR faces, after MEG recording in experiment 6, participants were shown Asian and white faces with neutral or painful expressions, which were taken from our previous work[13]. Faces with neutral expressions were used during MEG recording in experiment 6. Each face was presented in the centre of a computer monitor until a participant made a response. Participants were asked to rate the intensity of the pain portrayed by each face on a 9-point Likert scale (1 represents not painful at all and 9 represents extremely painful). Asian and white faces were presented in a random order.

Altruistic intention rating task. To estimate the intentions of participants to help OR and SR individuals suffering from physical pain, after MEG recording in experiment 6, participants were shown 48 video clips showing six Chinese models and six white models (half male and half female), which were taken from our previous work[54]. Each video lasted for 3 s and depicted a face with neutral expressions after receiving non-painful stimulation (a cotton swab touch) or with painful expressions after receiving painful stimulation (needle penetration) applied to the left or right cheeks of the face. Half of the video clips showed painful stimulations and half showed non-painful stimulations. After viewing the video clips, participants were required to rate to what extent they intended to help the models in the video clips using a Likert-type scale, where 1 indicated no effect and 7 indicated maximal effect (for example, 1 represents not at all and 7 represents extremely). Video clips showing Chinese and white models were presented in a random order. Difference scores (white(pain neutral) minus Asian(pain neutral)) were calculated to measure racial biases in empathy and altruistic intention.

IAT method. To evaluate the implicit attitudes of participants towards SR and OR faces, after the MEG recording in experiment 6, participants were asked to complete a race version of the IAT[52], in which they categorized SR faces/positive words with one key and OR faces/negative words with another key in two blocks and SR faces/negative words with one key and OR faces/positive words with another key in another two blocks. Response latency differences between the blocks with different response associations between faces and words reflected the relative ease of making associations between SR versus OR faces and concepts of good and bad. According to the established algorithm of the latencies[53], a positive IAT *D* score indicates that SR faces compared with OR faces are associated with good rather than bad, while a negative IAT *D* score indicates that SR faces compared with OR faces are associated with bad rather than good.

Procedures of brain imaging experiments. *EEG experiments.* For each trial, a face was displayed for 200 ms in the centre of a grey background, which was followed by a fixation cross with a duration varying randomly from 250 to 550 ms. Participants were asked to respond to the immediate repetition of the same face in two successive trials by pressing a button (one-back task). Each participant completed two EEG recording sessions. There were four runs for each EEG recording session. Each run consisted of 8 blocks of 18 trials (including 16 non-target faces and 2 target faces). There was a break of 8 s between the two consecutive blocks in which a number (that is, 8) at the fixation position flashed with a 1-s step and decreased to 1. Asian and black faces were used in the Asian/black sessions. Asian and white faces were used in the Asian/white sessions. The order of the two sessions was counterbalanced across participants. There were four runs in each session. In each run, faces were presented in the Rep-Cond in four blocks (two blocks for SR faces and two blocks for OR faces) and in the Alt-Cond in four blocks. Faces from one racial group were presented in the Rep-Cond and faces from two racial groups were presented the Alt-Cond. Faces were presented in a random order in each block. This design allowed the same number (that is, 128) of non-target trials of each racial category in the Alt-Cond and Rep-Cond. These were the same in all EEG experiments except that faces were presented upside-down in experiment 4 and that there were white/black sessions and white/Asian sessions for white participants in experiment 3. In addition, in experiment 2, each run consisted of 8 blocks of 18 trials in the Rep-Cond (2 blocks for female SR faces, 2 blocks for male SR faces, 2 blocks for female OR faces and 2 blocks for male OR-faces). There were 8 blocks of 18 trials in the Alt-Cond (4 blocks for female faces and 4 blocks for male faces). There were 128 non-target trials in each condition.

fMRI experiment. In experiment 5a, the face stimuli were projected onto a screen at the head of the magnet bore. Participants viewed the screen through a mirror attached to the head coil. For each trial, a face stimulus was displayed for 400 ms in the centre of a grey background, which was followed by a fixation cross with a duration varying randomly among 2, 4 and 6 s. Similarly, participants performed a one-back task that required responding to the immediate repetition of the same face in two successive trials by pressing a button. The fMRI experiment consisted of four runs. There were 9 blocks of 17 (or 18) trials (16 non-target and 1 or 2 target faces) in each run. In each run, faces were presented in the Rep-Cond in three blocks (with Asian, black or white faces in each block) and in the Alt-Cond in six blocks (Asian and black faces in two blocks, Asian and white faces in two blocks, and black and white faces in two blocks). Including an Alt-Cond with black and white faces would provide additional trials for examining RS effects. There was a 12.5-s break between two consecutive blocks, in which a green fixation was presented. These same procedures were used for experiment 5b, except that there were four blocks in each run, which consisted of two blocks in the Rep-Cond (with Asian or white faces in each block) and two blocks in the Alt-Cond (Asian and white faces). Moreover, a shorter duration (250 ms) and more variation of interstimulus intervals (1, 2, 3, 4, 5, s) were used.

MEG experiment. The procedures for MEG signal recording in experiment 6 were the same as those for experiment 5, except for the following. For each trial, a face stimulus was displayed for 200 ms followed by a fixation cross with a duration varying randomly from 250 to 550 ms. There were always 2 targets in each block and 128 non-target trials for each condition. After MEG recording, participants were asked to complete the old/new recognition task, the empathy rating task, the altruistic intention rating task and the IAT.

EEG data acquisition and analysis. A NeuroScan system (CURRY 7, Compumedics Neuroscan) was used for EEG recording and analysis for experiments 1a, 1b, 3 and 4. The EEG trace was continuously recorded from 32 scalp electrodes and was re-referenced to the average of the left and right mastoid electrodes offline. Impedances of individual electrodes were kept below 5 kΩ. Eye blinks and vertical eye movements were monitored using electrodes located above and below the left eye. The horizontal electro-oculogram was recorded from electrodes placed 1.5-cm lateral to the left and right external canthi. The EEG trace was digitized at a sampling rate of 1,000 Hz and subjected to an online band-pass filter of 0.01–400 Hz. EEG data were filtered with a low-pass filter at 30 Hz offline. Artefacts related to eye movement or eye blinks were removed using the covariance analysis tool implemented in CURRY 7 (ref. [79]). In experiment 2, the EEG trace was recorded using a Brain Products system (BrainAmp DC, Brain Products) using a band-pass filter of 0.1–100 Hz and digitized with a sampling rate of 500 Hz and referenced online against FCz; AFz was used as the ground electrode. The impedance of each electrode was kept below 5 kΩ. All electrodes were re-referenced to the left and right TP9/TP10 electrodes during offline processing. EEG data were analysed using Brain Vision Analyzer2 (Brain Products) offline and were filtered with a low-pass filter at 30 Hz. Artefacts related to eye movement or eye blinks were removed using the semiautomatic ICA application in Analyzer2 (Brain Products).

ERPs in each condition were averaged separately offline with an epoch beginning 200 ms before stimulus onset and continuing for 600 ms. Trials contaminated by eye movements and muscle potentials exceeding ±100 μV at any electrode were excluded from the average. Only EEG signals to non-target faces (128 trials in each condition) were included for analyses. Artefact rejection resulted in the following number of trials for each condition for further analyses: 109 ± 14 for experiment 1a; 101 ± 22 for experiment 1b; 118 ± 11 (Asian/black sessions) and 115 ± 13 (Asian/white sessions) for experiment 2; 108 ± 17 for experiment 3; and 110 ± 18 for experiment 4. The baseline for ERP measurements was the mean voltage of a 200-ms prestimulus interval and the latency was measured relative to the stimulus onset. Analyses of peak latencies of ERP components did not show any significant effect of RS and so were not reported in the main text. ERPs to non-target faces were characterized by an early negative wave at 95–125 ms (N1), a subsequent positive wave at 140–200 ms (P2) and a negative wave at 200–400 ms (N2) over the frontal/central electrodes (at 105–135 ms for N1 and 150–210 ms for P2 in experiment 3, which were delayed due to face inversion). To avoid potential significant but bogus effects on ERP amplitudes due to multiple comparisons[80], the mean values of the amplitudes of the N1, P2 and N2 components were calculated at frontal (Fz, FCZ, F3, F4, FC3 and FC4) and central (Cz, C3 and C4) electrodes. The mean amplitudes were then subject to repeated-measures ANOVA with race (SR versus OR) and condition (Rep-Cond versus Alt-Cond) as within-subject variables. RS effects were defined as decreased amplitudes to faces of a racial category in the Rep-Cond versus the Alt-Cond. Note that non-target faces in the Rep-Cond and Alt-Cond were identical across all participants. This design reduced the effect of low-level bottom-up processing of sensory and perceptual features of faces of different racial categories to a minimum.

ERPs to non-target faces also elicited a negative going wave at 140–200 ms (N170) (150–210 ms for inverted faces) over bilateral occipito-temporal electrodes (T5 and T6). Similarly, mean N170 amplitudes were subjected to ANOVAs with race (SR versus OR) and condition (Rep-Cond versus Alt-Cond) as within-subject variables.

fMRI data acquisition and analyses. Image acquisition was conducted using a GE 3.0 T MR scanner (HDx, Signa MR 750 System, GE Healthcare) with a standard head coil at the Center for MRI Research, Peking University. Functional images were acquired using T2-weighted, gradient-echo, echo-planar imaging sequences sensitive to BOLD contrast (64×64 matrix, 33 slices, 3.50×3.50×5.00 mm³ voxel; repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, field of view (FOV) = 22.4×22.4 cm, flip angle (FA) = 90°). A high-resolution anatomical T1-weighted image was acquired for each participant (256×256 matrix, 192 slices, 1.00×1.00×1.00 mm voxel; TR = 6.70 ms, TE = Min Full, FOV = 25.6×25.6 cm, FA = 12°). Padded clamps were used to minimize head motion and earplugs were used to attenuate scanner noises.

Functional images were preprocessed using the software SPM12 (the Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm). Functional scans were first corrected for within-scan acquisition time differences between slices and then realigned to the first volume to correct for inter-scan head motions. This realigning step provided a record of head motions within each fMRI run. Head movements were corrected within each run and six movement parameters (translation; x, y, z and rotation; pitch, roll, yaw) were extracted for further analysis in the statistical model. The functional images were resampled to 3×3×3 mm³ voxels, normalized to the MNI space using the parameters of anatomical normalization and then spatially smoothed using an isotropic of 8 mm full-width half-maximum (FWHM) Gaussian kernel.

Whole-brain analyses were conducted to examine RS effects on BOLD responses to Asian, black and white faces. The first level general linear model analyses modelled the onsets and durations of each stimulus for each participant. All conditions (race (Asian, white, black) × condition (Rep-Cond, Alt-Cond) and targets from all racial categories) were modelled for each participant. The fixed-effect model was used to estimate a canonical haemodynamic response function and its time derivatives. RS effects were defined as decreased activity to faces of a racial category in the Rep-Cond compared with the Alt-Cond. Random-effect analyses were then conducted based on statistical parameter maps from each participant to allow population inference. Significant activations were defined in the whole-brain analyses using a combined a voxel-level threshold of $P < 0.001$ and a cluster-level threshold of $P < 0.05$, FWE-corrected, according to recent suggestions[81]. MNI coordinates of activations are reported. Time courses of BOLD signals were extracted from ROIs using MarsBar Toolbox to illustrate repetition of neural responses to faces of a specific racial category. The software BrainNet was used to visualize our whole-brain results[82].

MEG data acquisition and analyses. *MEG and structural MRI data acquisition.* Cortical neuromagnetic activity was recorded using a whole-head MEG system with 102 magnetometers and 204 planar gradiometers (Elekta Neuromag TRIUX) in a magnetically shielded room. The MEG signals were sampled at 1 kHz with an online 0.1–330 Hz band-pass filter. Structural MRI of all the subjects' heads were collected using a 3.0 T GE Signa MR750 scanner (GE Healthcare). A three-dimensional magnetization-prepared rapid gradient echo T1-weighted sequence was used to obtain 1×1×1 mm³ resolution anatomical images. To co-register the MEG data with MRI coordinates, 3 anatomical landmarks (nasion, left and right pre-auricular points), 4 HPI coils and at least 200 points on the scalp and face were digitized using the Probe Position Identification system (Polhemus). The software Maxfilter (Elekta-Neuromag) with temporal signal space separation was first used to remove external interference from the raw MEG data. For each condition, the head positions of each pair of blocks were co-registered in reference to the position of the first block using Maxmove (a subcomponent of Maxfilter) before combining the MEG data.

Sensor-space whole-brain analysis. The offline analysis of MEG data was performed using Brainstorm[83]. Continuous MEG data were low-pass filtered at 30 Hz. Eye blink artefacts were attenuated with signal space projection by visually inspecting and removing the corresponding signal space projection component. The data were then epoched in accordance with stimulus trigger codes. Each epoch started 200 ms before face onset and continued for 600 ms. All epochs in which MEG activity exceeded 3,500 fT were removed from further analyses, resulting in the inclusion of at least 89 trials for each condition. The response to each face was baseline-corrected based on the 200-ms period preceding the face onset for each sensor.

MEG signals to OR faces were obtained by combining signals for black and white faces. MEG signals for Asian faces in Asian/black and Asian/white sessions were not combined because our EEG and fMRI results suggested that the processing of SR faces was influenced by contexts. A whole-brain cluster-based permutation t-test[84] was used to detect significant RS effects for faces of each racial category (Asian, black or white) on magnetometer signals. For faces of each racial category, a t-value was computed between MEG signals to faces in the Rep-Cond and the Alt-Cond. Adjacent points in time and space exceeding a predefined threshold ($P < 0.025$, two-tailed) were grouped into one or multiple clusters. The summed cluster t-values were compared against a permutation distribution that was generated by randomly reassigning condition membership for each participant (10,000 iterations) and computing the maximum cluster mass for each iteration. This approach reliably controls for multiple comparisons at the cluster level.

The time windows in which there were significant differences between the Rep-Cond and the Alt-Cond were used to guide subsequent source analyses. The permutation tests were performed at 100–200 ms, 200–300 ms and 300–400 ms based on the time courses of RS effects observed in the ERP results in experiments 1–3. We performed cluster correction in separate candidate windows to test hypotheses of early RS of neural responses to OR faces before 200 ms and late RS of neural responses to SR faces after 200 ms. We adopted windows of the same duration (that is, 100 ms) for cluster correction before and after 200 ms to avoid any potential artefacts due to cluster correction in time windows with different durations. As there were significant clusters at both 200–300 and 300–400 ms for black faces, we further combined these two time windows to characterize RS effects on neural responses to black faces. The RS effect at the negative clusters was defined as the opposite value of the Alt-Cond minus the Rep-Cond. The significant clusters were defined using a cluster-level threshold of $P < 0.05$ with corrections for multiple comparisons.

Source-space whole brain analysis. An anatomical T1 scan was acquired for each participant for source constructions of MEG signals. T1 images were obtained using a GE 3.0 T MR scanner (HDx, Signa MR 750 System, GE Healthcare). Segmentation of T1 images was conducted using automated algorithms provided in the software package FreeSurfer[85] (http://surfer.nmr.mgh.harvard.edu/).

Source localization and surface visualization were performed using the toolbox BrainStorm[83]. After co-registration between the individual anatomy and MEG sensors, cortical currents were estimated using a distributed model consisting of 15,002 current dipoles from the combined time series of magnetometer and gradiometer signals using a linear inverse estimator (weighted minimum-norm current estimate, signal-to-noise ratio of 3, depth weighting of 0.5) separately for each condition and for each participant in a single-sphere head model. Dipole orientations were unconstrained to the individual MRI scans. Noise covariance matrix was acquired from 2 min of empty-room MEG recordings collected daily before the experiment. For the group analysis, individual source-space data were projected to a standard brain model (Colin27, 15,002 vertices) using BrainStrorm[83]. For each of the 15,002 vertices, normalized source activations were obtained by computing the norm of three dipole moments in each direction and standardizing those values to pre-stimulus intervals of 200 ms (subtracting the mean and dividing by the standard deviation of the baseline). We also applied spatial smoothing using a 3-mm FWHM Gaussian kernel. For the purposes of group statistical analysis in the source space, the activity of each vertex over the time windows, in which significant activations were observed in the sensor space, was averaged. Whole-brain analyses of significant RS effects (that is, decreased neural responses to faces in the Rep-Cond versus the Alt-Cond) in source-space similarly used the cluster-based permutation t-test performed on 15,002 vertices of the cortical surface model (using a predefined threshold of $P < 0.025$, one-tailed, 10,000 iterations). One-tailed tests were implemented here to examine decreased neural responses to faces in the Rep-Cond versus the Alt-Cond. The significant clusters were defined using a cluster-level threshold of $P < 0.05$ with corrections for multiple comparisons.

Source-space ROI analysis. Based on the fMRI results obtained in experiment 5, we further performed ROI analyses to examine RS effects on source-space signals. Similar to other studies[86,87], ROIs were defined based on our fMRI results and anatomically based regions. Each participant's MRI scans were parcellated into anatomically based regions spanning the cortex[85], based on the Desikan–Killiany gyral atlas[88], using Freesurfer. Based on our fMRI results from experiment 5, the left frontal regions (left superior frontal and left rostral middle frontal) and the PCC (bilateral posterior cingulate and bilateral isthmus cingulate; 1,564 vertices in total) were included as the ROI for the analysis of RS effects related to white faces. The PCC/precuneus (bilateral posterior cingulate, bilateral isthmus cingulate and precuneus; 989 vertices in total) was included as the ROI for the analysis of RS effects related to black faces. The LTPJ (left supramarginal, inferiorparietal and superiorparietal; 1,140 vertices in total) was included as the ROI for the analysis of RS effects related to Asian faces. Because our MEG results showed RS effects of neural responses to black faces (as OR faces for Chinese participants) at 140–200 ms in the bilateral fusiform gyrus and LATL, we ran additional ROI analyses of activity in the temporal lobe (bilateral fusiform gyrus, left inferior, medial and superior temporal gyrus; 1,435 vertices in total) for white faces (also as OR faces for Chinese participants). All ROIs in each analysis were included as a single mask. The vertices for each of these labels were defined according to the Desikan–Killiany Atlas (https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation)[88]. The cluster-based permutation test (one-tailed, cluster alpha $P = 0.025$, 10,000 iterations) was performed within each mask for black faces at 140–200 and 230–400 ms and for Asian faces at 350–390 ms, as identified in the sensor-space analyses. For white faces, the cluster-based permutation test was performed at 140–200 ms, and 200–300 and 300–400 ms for exploration. Pairwise differences within the cluster were computed using paired-samples t-tests and corrected with FDR over tests.

Source-space time course analysis. To examine the fine spatiotemporal patterns of RS effects on neural responses to faces of each racial category, we extracted source-space time courses from ROIs (50 vertices in each ROI) in which the mean brain activity in a specific time window showed reliable RS effects on source-space MEG signals (see Supplementary Table 8 for MNI coordinates of these ROIs). We compared neural responses to the Alt-Cond versus the Rep-Cond at 0–400 ms after face onset. We performed one-tailed cluster-based permutation t-test (that is, a cluster refers to a set of contiguous time points) to control the multiple comparisons across time points using an a priori cluster threshold of $P < 0.025$, 10,000 iterations. One-tailed t-tests were conducted to examine RS effects (that is, greater responses towards the Alt-Cond than the Rep-Cond). Cluster P values were FDR-corrected across regions.

GCA method. To explore the potential causal influences between the key nodes of the neural network underlying the categorization of faces of each racial category, we conducted GCA, using the toolbox MVGC[89], of the averaged time courses of the RS effect (that is, Alt-Cond minus Rep-Cond at 0–400 ms after face onset) in the regions in which the source-space signals showed significant RS. ROIs involved in the categorization of black faces included the bilateral fusiform, the LATL, the mPFC and the PCC. ROIs involved in the categorization of white faces included the LIFC, the LATL and the PCC.

Time series of source-space signals were downsampled to 250 Hz before entering into GCA to ensure a reasonable model order for autoregressive modelling[90]. To satisfy the covariance stationarity assumption and the requirement of zero-mean time series for GCA, source-space signals were preprocessed, including linear detrending and rescaling (subtraction of the temporal mean and division by the temporal standard deviation), to remove drifts and slow fluctuations of source-space MEG signals and to normalize overall signal strengths of MEG signals in different brain regions. The optimal model order was selected using the Bayesian information criterion, resulting in a value of 4 (lag of 16 ms). We fitted a vector autoregression model to our time series data, using the Bayesian information criterion order as determined in the previous step, applying an ordinary least squares regression. After model estimation, we tested the stationarity of the model by examining whether the spectral radius $\rho(A)$ was larger than 1 for each participant. This analysis excluded one participant for further GCA of the time courses corresponding to black faces due to data violation of the stationarity assumption. To test the group-level significance of the averaged Granger causality, we generated 10,000 surrogate Granger causality matrices by shuffling the time series (that is, randomly permuting the time series of each ROI) to make the null distribution of the Granger causality. Significance of the original Granger causality matrix was tested by finding the position of original Granger causality in permutation distribution. Because the GCA engaged multiple brain regions and tested connections among these brain regions, we employed a stringent method for corrections of multiple comparisons (that is, Bonferroni correction, $P < 0.05$, as suggested by previous work[90,91].

To further assess the null hypothesis in experiments 1–3, we conducted Bayes factor analyses for repeated-measures ANOVA and paired t-tests. We calculated the Bayes factor in the program R v.3.5.1 (www.r-project.org) using the function anovaBF and ttestBF from the package BayesFactor[92]. We based the Bayes factor analyses on the default priors for ANOVA and paired t-test design (scale r on an effect size of 0.707). A Bayes factor indicates how much more likely each alternative model is supported compared with the null.

No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those reported in previous EEG[20] and MEG[86] studies. Data collection and analyses were not performed blinded to the conditions of the experiments. For statistical comparisons, data were assumed to be normal in distribution, but this was not formally tested. Individual participant data and distributions are represented in all relevant figures. All tests performed were two-tailed except the analysis of source-space MEG signals. One-tailed analyses were implemented in source-space MEG signals because the sensor-space signals had identified decreased amplitudes in the Rep-Cond compared with the Alt-Cond in the time windows in which we conducted further source-space analyses. The CIs are reported for effect sizes for ANOVA analyses (90% CIs of η_P^2), for the mean difference for t-tests and for the correlation coefficient for correlation analyses (95% CIs of mean difference or of the correlation coefficient).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability
The code used to analyse the data that support the findings of this study are available from the corresponding author upon reasonable request

References

1. Pager, D. & Shepherd, H. The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annu. Rev. Sociol.* **34**, 181–209 (2008).
2. Goyal, M. K., Kuppermann, N., Cleary, S. D., Teach, S. J. & Chamberlain, J. M. Racial disparities in pain management of children with appendicitis in emergency departments. *JAMA Pediatr.* **169**, 996–1002 (2015).
3. Anwar, S., Bayer, P. & Hjalmarsson, R. The impact of jury race in criminal trials. *Q. J. Econ.* **127**, 1017–1055 (2012).
4. Wetts, R. & Willer, R. Privilege on the precipice: perceived racial status threats lead White Americans to oppose welfare programs. *Soc. Forces* **97**, 793–822 (2018).
5. Meissner, C. A. & Brigham, J. C. Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol. Pub. Policy Law* **7**, 3–35 (2001).
6. Baron, A. S. & Banaji, M. R. The development of implicit attitudes: evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol. Sci.* **17**, 53–58 (2006).
7. McConnell, A. R. & Leibold, J. M. Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *J. Exp. Soc. Psychol.* **37**, 435–442 (2001).
8. Xu, X., Zuo, X., Wang, X. & Han, S. Do you feel my pain? Racial group membership modulates empathic neural responses. *J. Neurosci.* **29**, 8525–8529 (2009).
9. Avenanti, A., Sirigu, A. & Aglioti, S. M. Racial bias reduces empathic sensorimotor resonance with other-race pain. *Curr. Biol.* **20**, 1018–1022 (2010).
10. Han, S. Neurocognitive basis of racial ingroup bias in empathy. *Trends Cogn. Sci.* **22**, 400–421 (2018).
11. Ge, L. et al. Two faces of the other-race effect: recognition and categorisation of Caucasian and Chinese faces. *Perception* **38**, 1199–1210 (2009).
12. Qian, M. K. et al. Implicit racial biases in preschool children and adults from Asia and Africa. *Child Dev.* **87**, 285–296 (2016).
13. Sheng, F. & Han, S. Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *NeuroImage* **61**, 786–797 (2012).
14. Sheng, F., Han, X. & Han, S. Dissociated neural representations of pain expressions of different races. *Cereb. Cortex* **26**, 1221–1233 (2016).
15. Cosmides, L., Tooby, J. & Kurzban, R. Perceptions of race. *Trends Cogn. Sci.* **7**, 173–179 (2003).
16. Kawakami, K., Amodio, D. M. & Hugenberg, K. Intergroup perception and cognition: an integrative framework for understanding the causes and consequences of social categorization. *Adv. Exp. Soc. Psychol.* **55**, 1–80 (2017).
17. Gallagher, C. In *Theories of Race and Ethnicity* (eds Murji, K. & Solomos, J.) 40–56 (Cambridge Univ. Press, 2014).
18. Levin, D. T. Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *J. Exp. Psychol. Gen.* **129**, 559–574 (2000).
19. Hugenberg, K., Young, S. G., Bernstein, M. J. & Sacco, D. F. The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychol. Rev.* **117**, 1168–1187 (2010).
20. Ito, T. A. & Urland, G. R. Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *J. Pers. Soc. Psychol.* **85**, 616–626 (2003).
21. Ito, T. A. & Urland, G. R. The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. *Cogn. Affect. Behav. Neurosci.* **5**, 21–36 (2005).
22. Kubota, J. T. & Ito, T. A. Multiple cues in social perception: the time course of processing race and facial expression. *J. Exp. Soc. Psychol.* **43**, 738–752 (2007).
23. Ito, T. A. & Bartholow, B. D. The neural correlates of race. *Trends Cogn. Sci.* **13**, 524–531 (2009).
24. Wiese, H. The role of age and ethnic group in face recognition memory: ERP evidence from a combined own-age and own-race bias study. *Biol. Psychol.* **89**, 137–147 (2012).
25. Senholzi, K. B. & Ito, T. A. Structural face encoding: how task affects the N170's sensitivity to race. *Soc. Cogn. Affect. Neurosci.* **8**, 937–942 (2013).
26. Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I. & Bookheimer, S. Y. An fMRI investigation of race-related amygdala activity in African–American and Caucasian–American individuals. *Nat. Neurosci.* **8**, 720–722 (2005).
27. Phelps, E. A. et al. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* **12**, 729–738 (2000).
28. Golby, A. J., Gabrieli, J. D., Chiao, J. Y. & Eberhardt, J. L. Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* **4**, 845–450 (2001).
29. Richeson, J. A. et al. An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* **6**, 1323–1328 (2003).
30. Iidaka, T., Nogawa, J., Kansaku, K. & Sadato, N. Neural correlates involved in processing happy affect on same race faces. *J. Psychophysiol.* **22**, 91–99 (2008).
31. Cunningham, W. A. et al. Separable neural components in the processing of black and white faces. *Psychol. Sci.* **15**, 806–813 (2004).
32. Song, M. Introduction: who's at the bottom? Examining claims about racial hierarchy. *Ethnic Racial Stud.* **27**, 859–877 (2004).
33. Sadler, M. S., Correll, J., Park, B. & Judd, C. M. The world is not black and white: racial bias in the decision to shoot in a multiethnic context. *J. Soc. Issues* **68**, 286–313 (2012).
34. Gross, T. F. Own-ethnicity bias in the recognition of Black, East Asian, Hispanic, and White faces. *Basic Appl. Soc. Psych.* **31**, 128–135 (2009).
35. Young, A. W. Faces, people and the brain: the 45th Sir Frederic Bartlett Lecture. *Q. J. Exp. Psychol.* **71**, 569–594 (2018).
36. Kubota, J. T., Banaji, M. R. & Phelps, E. A. The neuroscience of race. *Nat. Neurosci.* **15**, 940–948 (2012).
37. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23 (2006).
38. Ringo, J. L. Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behav. Brain Res.* **76**, 191–197 (1996).
39. Vizioli, L., Rousselet, G. A. & Caldara, R. Neural repetition suppression to identity is abolished by other-race faces. *Proc. Natl Acad. Sci. USA* **107**, 20081–20086 (2010).
40. Heisz, J. J., Watter, S. & Shedden, J. M. Automatic face identity encoding at the N170. *Vis. Res.* **46**, 4604–4614 (2006).
41. Baillet, S. Magnetoencephalography for brain electrophysiology and imaging. *Nat. Neurosci.* **20**, 327–339 (2017).
42. Fiske, S. T. & Neuberg, S. L. A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. *Adv. Exp. Soc. Psychol.* **23**, 1–73 (1990).
43. Eberhardt, J. L., Goff, P. A., Purdie, V. J. & Davies, P. G. Seeing black: race, crime, and visual processing. *J. Pers. Soc. Psychol.* **87**, 876–893 (2004).
44. Plant, E. A., Goplen, J. & Kunstman, J. W. Selective responses to threat: the roles of race and gender in decisions to shoot. *Pers. Soc. Psychol. Bullet.* **37**, 1274–1281 (2011).
45. Wilson, J. P., Hugenberg, K. & Rule, N. O. Racial bias in judgments of physical size and formidability: from size to threat. *J. Pers. Soc. Psychol.* **113**, 59–80 (2017).
46. Ha, S. E. & Jang, S. J. Immigration, threat perception, and national identity: evidence from South Korea. *Int. J. Intercult. Relat.* **44**, 53–62 (2015).
47. Maddox, K. B. & Gray, S. A. Cognitive representations of Black Americans: reexploring the role of skin tone. *Pers. Soc. Psychol. Bullet.* **28**, 250–259 (2002).
48. Hayward, W. G., Rhodes, G. & Schwaninger, A. An own-race advantage for components as well as configurations in face recognition. *Cognition* **106**, 1017–1027 (2008).
49. Rossion, B. & Gauthier, I. How does the brain process upright and inverted faces? *Behav. Cogn. Neurosci. Rev.* **1**, 63–75 (2002).
50. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
51. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn (Lawrence Erlbaum Associates, 1988).
52. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
53. Greenwald, A. G., Nosek, B. A. & Banaji, M. R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197–216 (2003).
54. Han, S. et al. Empathic neural responses to others' pain are modulated by emotional contexts. *Hum. Brain Mapp.* **30**, 3227–3237 (2009).
55. Barth, F. *Ethnic Groups and Boundaries: The Social Organization of Culture Difference* (Waveland Press, 1998).
56. Quinn, P. C., Lee, K. & Pascalis, O. Perception of face race by infants: five developmental changes. *Child Dev. Perspect.* **12**, 204–209 (2018).
57. Byatt, G. & Rhodes, G. Identification of own-race and other-race faces: implications for the representation of race in face space. *Psychol. Bull. Rev.* **11**, 735–741 (2004).
58. Stolier, R. M. & Freeman, J. B. Functional and temporal considerations for top-down influences in social perception. *Psychol. Inq.* **27**, 352–357 (2016).
59. Freeman, J. B. & Johnson, K. L. More than meets the eye: split-second social perception. *Trends Cogn. Sci.* **20**, 362–374 (2016).
60. Kanwisher, N. & Yovel, G. The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 2109–2128 (2006).
61. Rice, G. E., Lambon Ralph, M. A. & Hoffman, P. The roles of left versus right anterior temporal lobes in conceptual knowledge: an ALE meta-analysis of 97 functional neuroimaging studies. *Cereb. Cortex* **25**, 4374–4391 (2015).

62. Mitchell, J. P., Heatherton, T. F. & Macrae, C. N. Distinct neural systems subserve person and object knowledge. *Proc. Natl Acad. Sci. USA* **99**, 15238–15243 (2002).

63. Amodio, D. M. The neuroscience of prejudice and stereotyping. *Nat. Rev. Neurosci.* **15**, 670–682 (2014).

64. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).

65. Gilbert, S. J., Swencionis, J. K. & Amodio, D. M. Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia* **50**, 3600–3611 (2012).

66. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).

67. Van der Cruyssen, L., Heleven, E., Ma, N., Vandekerckhove, M. & Van Overwalle, F. Distinct neural correlates of social categories and personality traits. *NeuroImage* **104**, 336–346 (2015).

68. Correll, J., Park, B., Judd, C. M. & Wittenbrink, B. The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *J. Pers. Soc. Psychol.* **83**, 1314–1329 (2002).

69. Wiese, H., Kaufmann, J. M. & Schweinberger, S. R. The neural signature of the own-race bias: evidence from event-related potentials. *Cereb. Cortex* **24**, 826–835 (2012).

70. Michel, C., Caldara, R. & Rossion, B. Same-race faces are perceived more holistically than other-race faces. *Vis. Cog.* **14**, 55–73 (2006).

71. Hughes, B. L. et al. Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception. *Proc. Natl Acad. Sci. USA* **116**, 14532–14537 (2019).

72. Nestor, A., Plaut, D. C. & Behrmann, M. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl Acad. Sci. USA* **108**, 9998–10003 (2011).

73. Freeman, J. B., Schiller, D., Rule, N. O. & Ambady, N. The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Hum. Brain Mapp.* **31**, 150–159 (2010).

74. Stolier, R. M. & Freeman, J. B. Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* **19**, 795–797 (2016).

75. Strohminger, N. et al. The MR2: a multi-racial, mega-resolution database of facial stimuli. *Behav. Res. Methods* **48**, 1197–1204 (2016).

76. Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res Methods* **47**, 1122–1135 (2015).

77. Valentine, T. & Endo, M. Towards an exemplar model of face processing: the effects of race and distinctiveness. *Q. J. Exp. Psychol.* **44**, 671–703 (1992).

78. Johnston, R. A., Kanazawa, M., Kato, T. & Oda, M. Exploring the structure of multidimensional face-space: the effects of age and gender. *Vis. Cogn.* **4**, 39–57 (1997).

79. Semlitsch, H. V., Anderer, P., Schuster, P. & Presslich, O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* **23**, 695–703 (1986).

80. Luck, S. J. & Gaspelin, N. How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* **54**, 146–157 (2017).

81. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl Acad. Sci. USA* **113**, 7900–7905 (2016).

82. Xia, M., Wang, J. & He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One* **8**, e68910 (2013).

83. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intel. Neurosci.* **8**, 879716 (2011).

84. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

85. Blanco-Elorrieta, E., Emmorey, K. & Pylkkänen, L. Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proc. Natl Acad. Sci. USA* **115**, 9708–9713 (2018).

86. Blanco-Elorrieta, E. & Pylkkänen, L. Bilingual language switching in the laboratory versus in the wild: the spatiotemporal dynamics of adaptive language control. *J. Neurosci.* **37**, 9022–9036 (2017).

87. Fischl, B. et al. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).

88. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

89. Barnett, L. & Seth, A. K. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* **223**, 50–68 (2014).

90. Seth, A. K. A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* **186**, 262–273 (2010).

91. Iwasaki, M., Noguchi, Y. & Kakigi, R. Neural correlates of time distortion in a preaction period. *Hum. Brain Mapp.* **40**, 804–817 (2019).

92. Morey, R. D., & Rouder, J. N. BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.11-11 (2015).

Acknowledgements

Author contributions

S.H. and Y.Z. conceived the research programme and designed the experiments. Y.Z., T.G., T.Z., W.L., T.W., X.H. and S.H. carried out the experiments. Y.Z., X.H. and S.H. analysed the data. S.H. and Y.Z. wrote the paper. S.H. supervised the entire work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information Extended data is available for this paper at https://doi.org/10.1038/s41562-019-0743-y.

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0743-y.

Correspondence and requests for materials should be addressed to S.H.

Peer review information Primary handling editor: Mary Elizabeth Sutherland.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Component	Session	Statistic value	ANOVA			Simple effect (Race)	
			Race	Condition	Race*Condition	Other-Race	Same-Race
P2 (140-200ms)	Asian/Black	F	69.29	21.17	19.85	31.56	0.32
		p	0.000	0.000	0.000	0.000	0.576
		η^2_p	0.652	0.364	0.349	0.460	0.009
		90% CI	[0.48 0.74]	[0.16 0.52]	[0.15 0.50]	[0.25 0.59]	[0 0.11]
	Asian/White	F	17.51	5.96	9.74	13.50	0.012
		p	0.000	0.020	0.003	0.001	0.912
		η^2_p	0.321	0.139	0.208	0.267	0.000
		90% CI	[0.12 0.48]	[0.01 0.30]	[0.04 0.38]	[0.08 0.43]	[0 0.03]
N2 (200-400ms)	Asian/Black	F	27.07	0.157	26.44	17.662	18.555
		p	0.000	0.704	0.000	0.000	0.000
		η^2_p	0.422	0.004	0.417	0.323	0.334
		90% CI	[0.21 0.56]	[0 0.09]	[0.21 0.56]	[0.12 0.48]	[0.13 0.49]
	Asian/White	F	18.87	1.30	8.86	8.62	1.98
		p	0.000	0.261	0.005	0.006	0.167
		η^2_p	0.338	0.034	0.193	0.189	0.051
		90% CI	[0.14 0.49]	[0 0.17]	[0.04 0.36]	[0.03 0.36]	[0 0.19]

Extended Data Fig. 1 | Statistical results of P2 and N2 amplitudes in experiment 1A. [a]Note: Effect size is indexed as the partial eta-squared value. The 90% CIs are reported for partial eta-squared values.

Component	Session	Statistic value	ANOVA			Simple effect (Race)	
			Race	Condition	Race*Condition	Other-Race	Same-Race
P2 (140-200ms)	Asian/Black	F	27.91	17.27	39.83	38.74	1.12
		p	0.000	0.000	0.000	0.000	0.296
		η^2_p	0.430	0.318	0.518	0.511	0.029
		90% CI	[0.22 0.57]	[0.12 0.48]	[0.32 0.64]	[0.31 0.63]	[0 0.16]
	Asian/White	F	2.58	8.54	9.86	14.09	0.007
		p	0.117	0.006	0.003	0.001	0.932
		η^2_p	0.065	0.188	0.210	0.276	0.000
		90% CI	[0 0.22]	[0.03 0.36]	[0.05 0.38]	[0.09 0.44]	[0 0.01]
N2 (200-270ms)	Asian/Black	F	10.60	7.44	50.47	40.76	6.04
		p	0.002	0.010	0.000	0.000	0.019
		η^2_p	0.223	0.167	0.577	0.524	0.140
		90% CI	[0.05 0.39]	[0.02 0.34]	[0.38 0.68]	[0.32 0.64]	[0.01 0.31]
	Asian/White	F	6.47	6.06	16.57	13.222	2.196
		p	0.015	0.019	0.000	0.001	0.147
		η^2_p	0.149	0.141	0.309	0.263	0.056
		90% CI	[0.02 0.32]	[0.01 0.31]	[0.11 0.47]	[0.08 0.43]	[0 0.20]

Extended Data Fig. 2 | Statistical results of P2 and N2 amplitudes in experiment 1B. [a]Note: Effect size is indexed as the partial eta-squared value. The 90% CIs are reported for partial eta-squared values.

	ANOVA	Asian/Black				Asian/White			
		F	p	η_p^2	90% CI	F	p	η_p^2	90% CI
P2 (160 – 220 ms)	Race	76.04	<0.001	0.685	[0.52 0.76]	144.59	<0.001	0.828	[0.71 0.88]
	FG	16.06	<0.001	0.315	[0.11 0.48]	0.172	0.681	0.006	[0 0.11]
	Condition	54.34	<0.001	0.608	[0.42 0.71]	18.02	<0.001	0.483	[0.15 0.54]
	Race× FG	0.237	0.630	0.007	[0 0.10]	0.280	0.600	0.009	[0 0.12]
	Race × Condition	23.45	<0.001	0.401	[0.19 0.55]	7.953	0.008	0.210	[0.03 0.39]
	FG× Condition	0.068	0.796	0.002	[0 0.072]	0.000	0.985	0.000	[0 0]
	Race×FG×Condition	2.703	0.109	0.072	[0 0.23]	2.882	0.100	0.088	[0 0.26]
N2 (220 – 400 ms)	Race	35.35	<0.001	0.503	[0.29 0.63]	94.23	<0.001	0.759	[0.61 0.83]
	FG	13.11	0.001	0.272	[0.08 0.44]	0.09	0.766	0.003	[0 0.09]
	Condition	0.162	0.690	0.005	[0 0.09]	4.94	0.034	0.141	[0.01 0.32]
	Race× FG	0.811	0.374	0.023	[0 0.15]	3.025	0.092	0.092	[0 0.27]
	Race × Condition	23.69	<0.001	0.404	[0.19 0.55]	16.37	<0.001	0.353	[0.13 0.52]
	FG× Condition	0.391	0.536	0.011	[0 0.12]	1.30	0.263	0.042	[0 0.20]
	Race× FG× Condition	0.554	0.462	0.016	[0 0.13]	1.90	0.178	0.060	[0 0.22]

Extended Data Fig. 3 | Statistical results of P2 and N2 amplitudes in experiment 2. [a]Note: Effect size is indexed as the partial eta-squared value. FG: Face gender. The reported for partial eta-squared values.

Compone nt	Session	Statistic value	ANOVA			Simple effect (Race)	
			Race	Condition	Race*Condition	Other-Race	Same-Race
P2 (140-200ms)	White/Black	F	21.65	7.84	31.56	32.23	3.37
		p	0.000	0.009	0.000	0.000	0.076
		η^2_p	0.411	0.202	0.504	0.510	0.098
		90% CI	[0.18 0.56]	[0.03 0.38]	[0.28 0.64]	[0.28 0.64]	[0 0.27]
	White/Asian	F	4.69	6.21	4.38	9.81	0.31
		p	0.038	0.018	0.044	0.004	0.582
		η^2_p	0.121	0.154	0.114	0.224	0.009
		90% CI	[0 0.29]	[0.02 0.33]	[0 0.28]	[0.05 0.40]	[0 0.12]
N2 (200-400ms)	White/Black	F	23.35	1.90	30.30	6.48	19.40
		p	0.000	0.178	0.000	0.016	0.000
		η^2_p	0.430	0.058	0.494	0.173	0.385
		90% CI	[0.20 0.58]	[0 0.22]	[0.27 0.63]	[0.02 0.35]	[0.16 0.54]
	White/Asian	F	32.66	0.369	6.79	3.34	2.69
		p	0.000	0.547	0.014	0.076	0.110
		η^2_p	0.490	0.011	0.166	0.090	0.073
		90% CI	[0.27 0.62]	[0 0.12]	[0.02 0.34]	[0 0.25]	[0 0.23]

Extended Data Fig. 4 | Statistical results of P2 and N2 amplitudes in experiment 3. [a]Note: Effect size is indexed as the partial eta-squared value. FG: Face gender. The reported for partial eta-squared values.

Component	Session	Statistic value	ANOVA			Simple effect (Race)	
			Race	Condition	Race*Condition	Other-Race	Same-Race
P2 (150–210ms)	Asian/Black	F	6.90	4.00	1.02	4.60	0.11
		p	0.013	0.054	0.319	0.039	0.743
		η^2_p	0.169	0.015	0.029	0.119	0.003
		90% CI	[0.02 0.34]	[0 0.27]	[0 0.16]	[0 0.29]	[0 0.09]
	Asian/White	F	4.03	11.01	0.31	6.57	4.51
		p	0.053	0.002	0.582	0.015	0.041
		η^2_p	0.106	0.245	0.009	0.162	0.117
		90% CI	[0 0.27]	[0.06 0.42]	[0 0.12]	[0.02 0.34]	[0 0.29]
N2 (210–410ms)	Asian/Black	F	5.91	7.19	6.95	18.50	0.13
		p	0.020	0.011	0.013	0.000	0.723
		η^2_p	0.148	0.175	0.170	0.352	0.004
		90% CI	[0.01 0.32]	[0.02 0.35]	[0.02 0.34]	[0.14 0.51]	[0 0.09]
	Asian/White	F	15.77	3.54	2.94	8.61	0.063
		p	0.000	0.069	0.095	0.006	0.803
		η^2_p	0.317	0.094	0.080	0.202	0.002
		90% CI	[0.11 0.48]	[0 0.26]	[0 0.24]	[0.04 0.38]	[0 0.072]

Extended Data Fig. 5 | Statistical results of P2 and N2 amplitudes in experiment 4. [a]Note: Effect size is indexed as the partial eta-squared value. FG: Face gender. The reported for partial eta-squared values.

A DCM model space
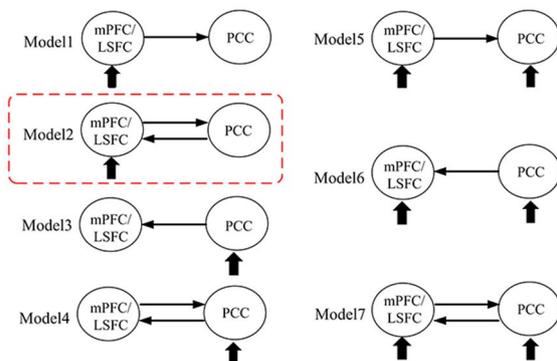
B Winning model
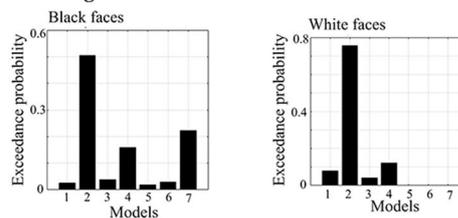
C Parameter estimation
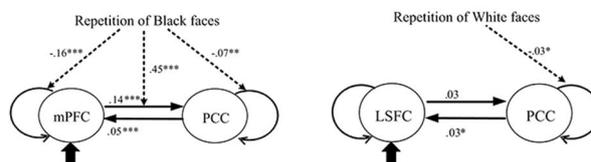
Extended Data Fig. 6 | Model structures and results of the DCM analysis in experiment 5a. (a) Illustration of individual models in the DCM model space. The models were different in the region that driving vision input were assigned to and the intrinsic connectivity between regions. Only the intrinsic connectivity between brain regions and the driving visual input were plotted. The modulation effect of on each intrinsic connectivity and the intrinsic connectivity within each regions of the single models were omitted. **(b)** The exceedance probabilities of single models of Black and White faces. Model 2 marked inside the red square had the highest exceedance probability for both Black and White faces. **(c)** Strength of the intrinsic and modulatory connectivity estimated based on Model 2 for Black and White faces, respectively. Repetition of Black faces significantly modulated within-region connectivity in both mPFC and PCC (N=53; mPFC to mPFC: t(52) = −3.06, p = 0.003, Cohen's D = 0.42, 95% CI = −0.27, −0.06, PCC to PCC : t(52) = −3.34, p = 0.002, Cohen's D = 0.46, 95% CI = −0.12, −0.03) and between-region connectivity from mPFC to PCC (t(52) = 6.12, p < 0.001, Cohen's D = 0.84, 95% CI = 0.30, 0.60). By contrast, repetition of White faces only significantly modulated within- region connectivity for PCC (t(52) = −2.10, p = 0.041, Cohen's D = 0.29, 95% CI = −0.06, −0.001). *p<0.05,** p<0.01,*** p<0.001.

	White faces			Asian faces			Black faces		
	Rep	Alt (Asian/White session)	Alt (Black/White session)	Rep	Alt (Asian/White session)	Alt (Asian/Black session)	Rep	Alt (Asian/Black session)	Alt (Black/White session)
RT(ms)	515.2±64.5	520.4±54.5	522.8±53.3	519.2±55.4	513.5±53.4	529.5±55.7	520.0±52.9	508.4±44.9	515.4±55.1
Accuracy (%)	83±12.6	82±21.9	79±15.8	84±12.6	82±16.2	80±15.1	81±20.0	81±16.1	82±16.0

Extended Data Fig. 7 | RTs and accuracies (mean ± SD) in experiment 6.
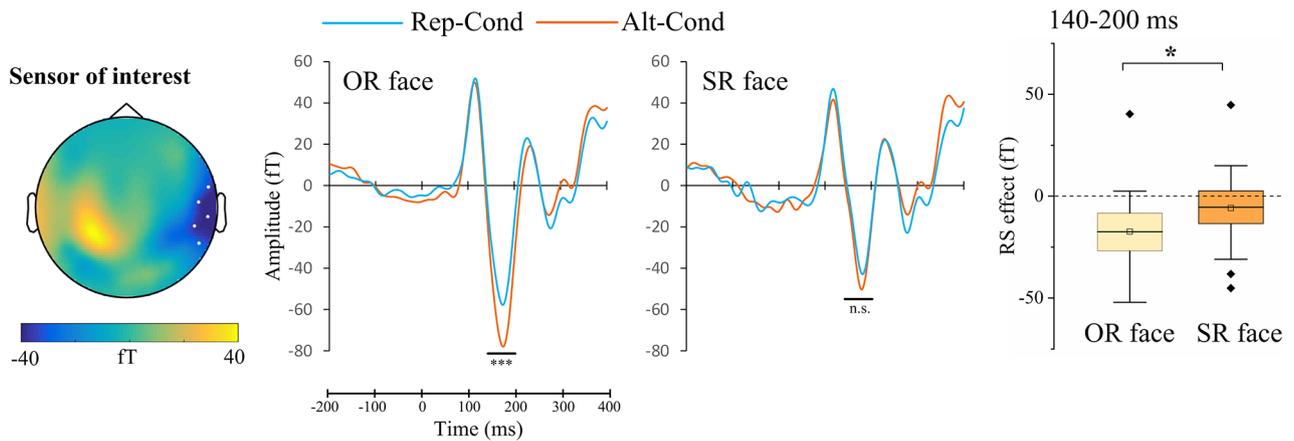
Extended Data Fig. 8 | Results of MEG sensor-space signals in experiment 6. The left panel shows the magnetometer sensors that showed the strongest neural responses at 140–200 ms across all conditions (p <0.05, FDR corrected). The middle panel shows mean MEG responses over these sensors to OR-faces and SR-faces. The right bar charts illustrate the mean RS effects in sensor-space MEG signals (Alt-Cond > Rep-Cond) by showing quartiles (boxes), means (square inside boxes), medians (horizontal lines inside boxes), maximum and minimum excluding outliers(whiskers), and outliers (diamonds). ANOVAs of the mean sensor-space magnetic responses at 140–200 ms confirmed a larger RS effect for OR-(collapsing Black and White faces) than SR- (Asian) faces (N=26; $F_{(1,25)} = 6.663$, p = 0.016, $\eta_p = 0.210$, 90% CI = 0.02, 0.41).
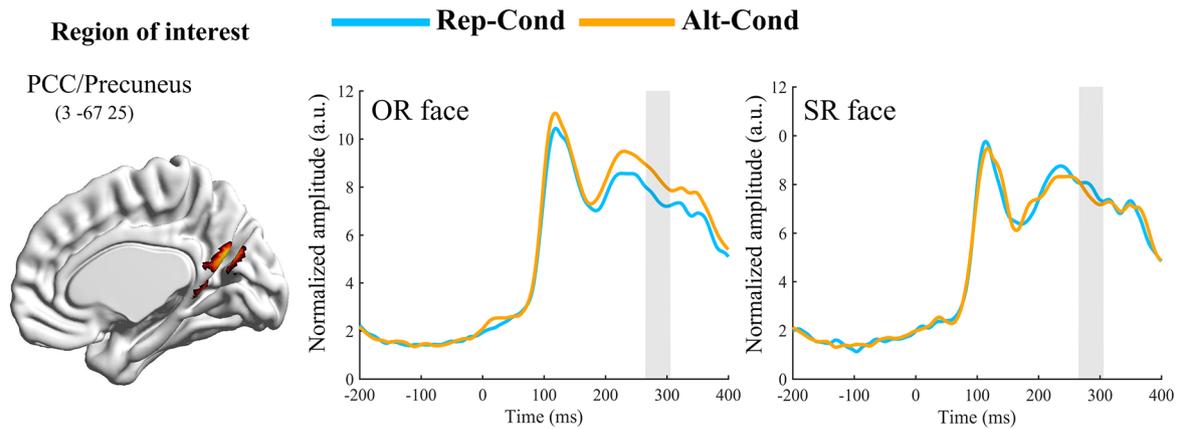
Extended Data Fig. 9 | MEG results of ROI analyses in experiment 6. Based on our fMRI results, we predicted stronger RS effects on source-space signals in PCC to OR- than SR-faces. To test this, we extracted time courses of MEG signals in the PCC (3/-67/25) identified in our fMRI results (shown in the left panel) and compared the RS effects (that is. Alt-Cond vs. Rep-Cond) between OR-faces and SR-faces at each time points. Cluster based permutation tests (one-tailed) were conducted across time points using a priori cluster threshold p < 0.025, 10,000 iterations. This test yielded a significant cluster showing greater RS effect for OR-faces than SR-faces at 266–305 ms in PCC (N=26; cluster p = 0.028), as illustrated in the middle and right panels. a.u.= arbitrary unit.
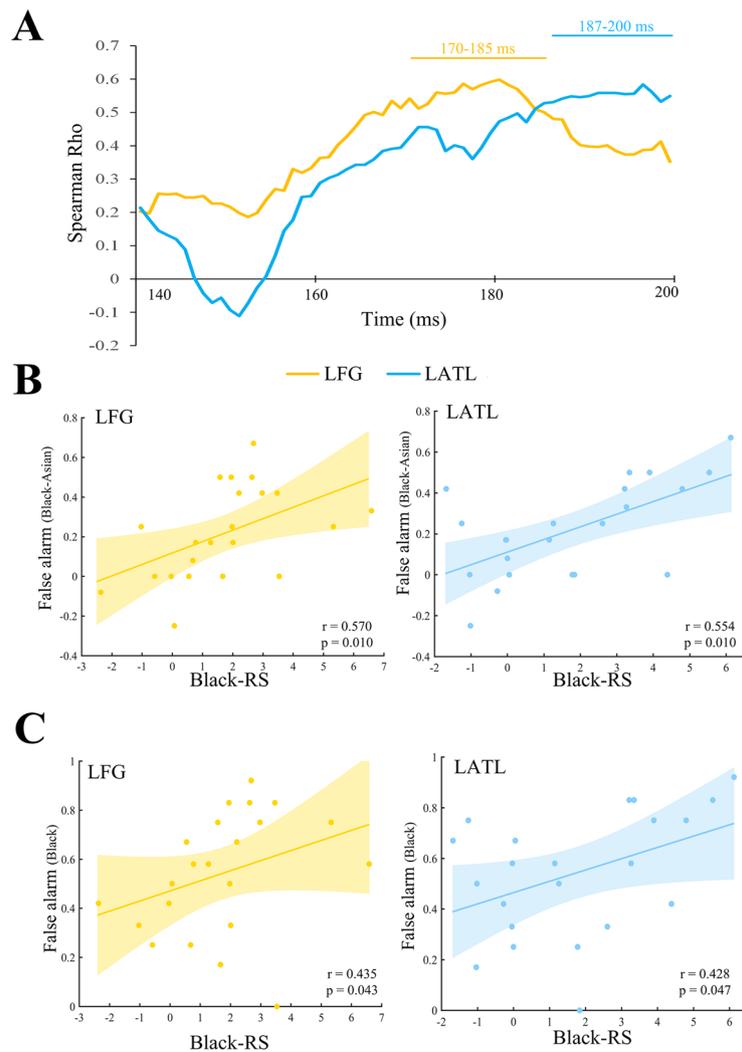
Extended Data Fig. 10 | Coupling between neural categorization of OR-faces and racial biases in cognition and altruistic intension in experiment 6.
(A)The time course of Spearman rank correlations between RS of LFG (yellow line) and LATL (blue line) activity and false alarm rates during face recognition. Correlations were calculated point-by-point at 140–200 ms. Significant correlations were observed at 170- 185 ms for LFG activity and at 187–200 ms for LATL activity (p < 0.05, FDR corrected). Averaged RS effects on LFG (170–185 ms) and LATL (187–200 ms) activities predicted false alarms during recognition of Black (vs. Asian) faces (N=22; LFG: r = 0.570, p = 0.010, 95% CI = 0.18, 0.83; LATL: r = 0.554, p = 0.010, 95% CI = 0.08, 0.87, FDR corrected). (**C**) Averaged RS effects on LFG (170–185 ms) and LATL (187- 200 ms) activities predicted false alarms during recognition of Black faces (LFG: r = 0.435, p = 0.043, 95% CI = 0, 0.76; LATL: r = 0.428, p = 0.047, 95% CI = −0.06, 0.75, uncorrected).

nature research

Corresponding author(s):	Shihui Han
Last updated by author(s):	Aug 20, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	A NeuroScan system (CURRY 7, Compumedics Neuroscan, Texas) was used for EEG recording. MRI data were collected using a GE 3.0 T MR scanner (HDx, Signa MR 750 System; GE Healthcare, Milwaukee, WI). MEG data were recorded using a whole-head MEG system with 102 magnetometers and 204 planar gradiometers (Elekta Neuromag TRIUX).
Data analysis	A NeuroScan system (CURRY 7, Compumedics Neuroscan, Texas) was used for EEG data analysis. fMRI data were processed using SPM12 software (the Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm). MEG data analysis was Brainstorm package (01-Nov-2018).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data available on request from the authors

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine the sample size. The sample size of the pilot experiment (n=57) was comparable with the previous study using a similar paradigm (e.g., Byatt & Rhodes, 2004).The sample sizes used in our EEG experiments (Exp. 2-4, n = 32-38) are comparable with recent EEG studies of racial bias (e.g., Ito & Urland, 2003). A large sample was recruited in Exp. 5 A (fMRI experiment, n=55) and Exp. 5B (n=40) because the long interstimulus intervals used might weaken RS effects on BOLD responses to faces. The sample size of Exp. 6 (MEG experiment) was comparable with the conventional sample size of current MEG studies (e.g., Blanco-Elorrieta & Pylkkänen, 2017).
Data exclusions	Two participants were excluded from fMRI data analyses in Experiment 5A due to head movement during scanning. Three participants failed EEG recording in White/Black session in Experiment 3. One participant was excluded from GCA of the time courses corresponding to Black faces due to data violation of the stationarity assumption in Experiment 6.
Replication	EEG results of Exp. 1 were replicated in Exp. 2 and 3 in different samples and using different sets of stimuli. The temporal profiles of EEG results were replicated by MEG results. fMRI results were replicated by MEG source-space signals.
Randomization	Different conditions in all experiments were presented in random orders. Stimuli in each condition were also displayed in a random order.
Blinding	No group allocation was used, and the investigators were blinded to stimulus conditions during EEG/fMRI/MEG data recording.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The population characteristics of participants were disclosed in methods.
Recruitment	Participants were recruited from volunteers of university students who were paid for their participation.
Ethics oversight	This study was approved by the local Research Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Task; event-related design
Design specifications	On each trial, a face stimulus was displayed for 400 ms in the center of a gray background which was followed by a

Design specifications	fixation cross with a duration varying randomly among 2, 4, and 6 s. Similarly, participants performed a one-back task that required responding to the immediate repetition of the same face in two successive trials by pressing a button. The fMRI experiment consisted of 4 runs. There were 9 blocks of 17 (or 18) trials (16 non-target and 1 or 2 target faces) in each run. In each run, faces were presented in Rep-Cond in 3 blocks (with Asian, Black, or White faces in each block) and in Alt-Cond in 6 blocks (Asian and Black faces in 2 blocks, Asian and White faces in 2 blocks, and Black and White faces in 2 blocks).
Behavioral performance measures	Response accuracies and reaction times were recorded during fMRI scanning. The mean and SD were tested to prove that participants attended to the task as expected.

Acquisition

Imaging type(s)	Functional and structural
Field strength	3 T
Sequence & imaging parameters	Functional images were acquired by using T2-weighted, gradient-echo, echo-planar imaging sequences sensitive to BOLD contrast [64 × 64 matrix, 33 slices, 3.50×3.50×5.00 mm3 voxel; repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, field of view (FOV) =22.4 × 22.4 cm, flip angle (FA) = 90°]. A high-resolution anatomical T1-weighted image was acquired for each participant (256 × 256 matrix, 192 slices, 1.00 × 1.00 × 1.00 mm voxel; TR = 6.70ms, TE Min Full, FOV = 25.6 × 25.6 cm, FA = 12°).
Area of acquisition	whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	SPM12 on MATLAB R2016b
Normalization	The standard normalization procedure in SPM12 was used.
Normalization template	MNI template
Noise and artifact removal	Motion-correction parameters were included into the GLM.
Volume censoring	We discarded volumes recorded at initial 6 sec in each run.

Statistical modeling & inference

Model type and settings	Univariate. The first-level analysis used a fixed effect model to estimate a canonical hemodynamic response function and its time derivatives. Random-effect analyses were then conducted based on statistical parameter maps from each participant to allow population inference.
Effect(s) tested	Mean value of repetition suppression of BOLD responses over the participants

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s) *Describe how anatomical locations were determined (e.g. specify whether automated labeling algorithms or probabilistic atlases were used).*

Statistic type for inference (See Eklund et al. 2016)	For the whole-brain analysis, we employed a cluster-level correction for multiple comparisons implemented in SPM12 (cluster-forming threshold P = 0.001, cluster level FWE corrected, p < 0.05).
Correction	FWE

Models & analysis

n/a | Involved in the study
☐ | ☒ Functional and/or effective connectivity
☒ | ☐ Graph analysis
☒ | ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	Dynamic causal modeling was used to assess the effective connectivity between brain regions that showed significant RS effects for Black and White faces and to examine how the effective connectivity was modulated by alternating vs. repetition conditions.