

Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses

Feng Sheng^a, Shihui Han^{a,b}

^a Department of Psychology, Peking University, PR China

^b Beijing Key Laboratory of Learning and Cognition, Department of Psychology, Capital Normal University, PR China

article info

Article history:

Accepted 11 April 2012

Available online 20 April 2012

Keywords:

Empathy

Event-related potential

Racial bias

Intergroup relationships

abstract

Social relationships affect empathy in humans such that empathic neural responses to perceived pain were stronger to racial in-group members than to racial out-group members. Why does the racial bias in empathy (RBE) occur and how can we reduce it? We hypothesized that perceiving an other-race person as a symbol of a racial group, rather than as an individual, decreases references to his/her personal situation and weakens empathy for that person. This hypothesis predicts that individuating other-race persons by increasing attention to each individual's feelings or enclosing other-race individuals within one's own social group can reduce the RBE by increasing empathic neural responses to other-race individuals. In Experiment 1, we recorded event related brain potentials from Chinese adults as they made race judgments on Asian and Caucasian faces with pain or neutral expressions. We identified the RBE by showing that, relative to neutral expressions, pain expressions increased neural responses at 128–188 ms after stimulus onset over the frontal/central brain regions, and this effect was evident for same-race faces but not for other-race faces. Experiments 2 and 3 found that paying attention to observed individual's feelings of pain and including other-race individuals in one's own team for competitions respectively eliminated the RBE by increasing neural responses to pain expressions in other-race faces. Our results indicate that the RBE is not inevitable and that manipulations of both cognitive strategies and intergroup relationships can decrease RBE-related brain activity.

© 2012 Elsevier Inc. All rights reserved.

Introduction

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others. . . Adam Smith in *The Theory of Moral Sentiments* (1759, p.1)

It is widely acknowledged that empathy benefits human society by promoting prosocial behaviors (de Waal, 2008). Recent behavioral studies suggest that empathy is affected by social group relationships. For example, white students reported greater feelings of empathy for and assigned more lenient punishments to a white than to a black defendant (Johnson et al., 2002), and pro-white empathy biases to patients' pain expressions predicted pro-white pain treatment biases (Drwecki et al., 2011). These findings indicate a racial bias in empathy

subjective reports, provides evidence that the human brain is able to understand others' feelings. Moreover, the brain activity elicited by perceived pain of a target person is overlapped with the brain activity underlying the first hand pain experience and is associated with an observer's own distress at witnessing another person's suffering. These suggest that an observer feels the same emotion that a target person feels or shares others' feelings.

Interestingly, recent research showed that empathy-related brain activity was modulated by racial intergroup relationships between an observer and a target. For example, Xu et al. (2009) scanned Chinese and Caucasian participants during viewing video clips showing faces of Chinese or Caucasian models who received painful (needle penetration) or non-painful (Q-tip touch) stimulation. They found that ACC activity to perceived pain decreased to racial out-group than to racial in-group members in both Chinese and Caucasian participants. Similarly, Avenanti et al. (2010) recorded sensorimotor brain responses from white-Caucasians and black-Africans who watched video clips depicting needle penetrating or Q-tip touching hands of stranger black or white models. They found that sensorimotor responses decreased to perceived painful stimulation applied to body parts of racial out-group versus in-group individuals in both white and black participants (Avenanti et al., 2010). Mathur et al. (2010) recorded BOLD responses from African-American and Caucasian-American individuals while they perceived naturalistic visual scenes depicting African-American or Caucasian-American individuals in painful (e.g., in the midst of a natural disaster) or neutral (e.g., attending an outdoor picnic) situations. They showed that African-American individuals additionally recruited the medial prefrontal cortex when observing the suffering of members of their own social group. Moreover, the increased medial prefrontal activity to pain expressed by racial in-group relative to out-group members predicted altruistic motivation for racial in-group members. These findings suggest that multiple levels of neural mechanisms involved in affective sharing and sensorimotor resonance are modulated by race-based social relationships and thus mediate the RBE shown in behaviors.

As the RBE is related to both within-group altruism and between-group conflict (Galinsky et al., 2011), discovering methods to reduce RBE-related brain activity has high social importance and may further our understanding of the neurocognitive mechanisms underlying the RBE. As other-race faces are perceived as more psychologically similar to each other relative to same-race faces (Valentine and Endo, 1992; Vizioli et al., 2010), an other-race face may be perceived as a symbol of a group rather than of an individual. This may lead an observer to perceive a racial out-group member without any reference to the individual's personal situation (Kinder and Sears, 1981) and result in decreased empathy. This hypothesis predicts that (1) increasing attention to an individual's painful feelings (a cognitive strategy that enhances the individuated processing of persons) or (2) including other-race individuals in one's own social group (which also leads to individuated processing of in-group members (Sporer, 2001)) will reduce the RBE by increasing empathic neural responses to other-race individuals.

Previous research has shown that enhancing attention to others' individual attributes (Wheeler and Fiske, 2005) and including other-race individuals in one's own team for competition (Van Bavel et al., 2008) reduced the racial bias in amygdala activity when observing neutral faces. Here we assessed if manipulations of attention and intergroup relationships can reduce RBE-related brain activity. In three experiments, we recorded ERPs from Chinese adults while they perceived Asian and Caucasian faces with neutral or pain expressions (Fig. 1A). Similar to the previous fMRI (Gu and Han, 2007; Gu et al., 2010; Han et al., 2009; Jackson et al., 2005; Saarela et al., 2007; Singer et al., 2004) and ERP (Decety et al., 2010; Fan and Han, 2008; Li and Han, 2010) studies, Experiment 1 in the current work aimed to associate the brain activity with empathy in two senses. First, we examined whether ERPs differentiate between pain and neutral expressions and whether the differential neural activity to pain versus neutral expressions has a source in the brain regions identified in the fMRI research (e.g., ACC). Second, we assessed whether the differential neural activity to pain versus neutral expressions is associated with subjective feelings of self-unpleasantness induced by the perceived pain expression. Moreover, Experiment 1 tested RBE by examining whether empathic neural responses (i.e., increased neural activity to pain than neutral expressions) during race judgments were stronger to same-race faces than other-race faces.

While the fMRI studies found that the ACC and medial prefrontal cortex were engaged in the RBE for pain (Mathur et al., 2010; Xu et al., 2009), the ERP studies have shown that empathic neural responses to others' pain varied dynamically as a function of the processing time (Decety et al., 2010; Fan and Han, 2008; Han et al., 2008; Li and Han, 2010). An early component of empathic neural responses occurred regardless of task demands whereas a late component of empathic neural responses was enhanced by task demands that required attention to others' pain (Fan and Han, 2008). There has been ERP evidence for sensitivity to race at the early perceptual processing. Vizioli et al. (2010) showed that repetitive presentations of same-race faces of same identity decreased the amplitude of the face-sensitive N170 component. However, such repetition suppression of the N170 occurred to other-race faces regardless of identity. They also found that inverted same-race faces led to greater recognition impairment and elicited larger N170 amplitudes compared to inverted other-race faces (Vizioli et al., 2011). Moreover, the race-sensitive N170 amplitude was affected by preexisting racial attitudes (Ofan et al., 2011). Ito and colleagues also found that an early frontal negativity around 120 ms increased to black than to white faces in Caucasians and a following positivity around 170 ms showed larger amplitude to black than to white faces and to emotional (angry and happy) than to neutral expressions (Ito and Urland, 2003, 2005; Kubota and Ito, 2007). While the current ERP findings suggest that neural encoding of same-/other-race faces occurs at early stages of processing over the anterior and posterior brain regions, to date, there has been no evidence that the early ERP responses to facial pain expressions are modulated by race. By recording ERPs to pain and neutral expression of racial in-group and out-group faces,

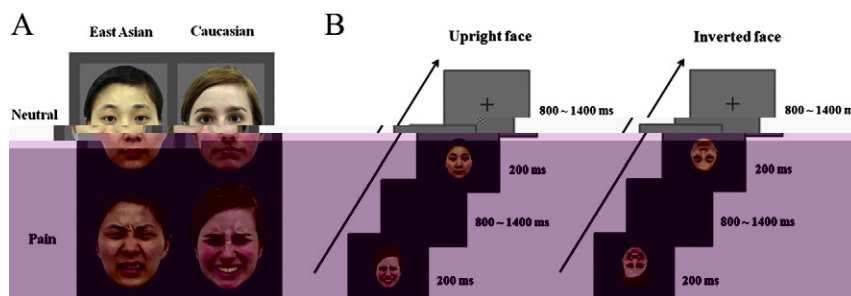


Fig. 1. Materials and procedures in Experiment 1. (A) Illustration of Asian and Caucasian faces with neutral and pain expressions used in our study. (B) Illustration of the EEG recording procedure. Upright and inverted faces were presented in separate blocks of trials.

Experiment 1 in the current study examined the time course of neural responses to pain versus neutral expressions and whether differential neural responses to pain versus neutral expressions were modulated by racial group relationship between observers and perceived targets. Participants in Experiment 1 were also asked to perform race judgments on a set of inverted Asian and Caucasian faces in order to control for effects produced by perceptual feature differences between Asian and Caucasian faces though the luminance level was matched for the two sets of stimuli.

Most importantly, Experiments 2 and 3 investigated whether enhanced attention to another individual's feelings of pain or including other-race individuals in one's own team during a competitive situation can eliminate RBE-related brain activity. Experiment 2 asked participants to perform both race judgments and pain judgments on racial in-group and out-group faces. If enhanced attention to an individual's feelings increases empathy for others' pain, we would expect that, relative to race judgments, pain judgments increase empathy-related neural activity to racial out-group faces and lead to reduction of the RBE. Experiment 3 manipulated intergroup relationships between participants and perceived faces. Participants were assigned to the blue or green team for a competitive game and both the fellow-team and opponent-team consisted of half Asians and half Caucasians. ERPs were recorded during race judgments on faces of fellow-team and opponent-team members. If in-group relationships increase empathy for other-race individuals of the fellow-team, we would expect increased empathy-related neural activity to Caucasian faces of the fellow-team and this may also lead to reduction of the RBE for Caucasian faces of the fellow-team compared to Caucasian faces of the opponent-team.

Materials and methods

Participants

Forty-eight Chinese college students were enrolled in the study as paid volunteers. There were 16 participants (half males) in each experiment (mean age \pm SD=23.81 \pm 1.47 year in Experiment 1; 21.75 \pm 1.88 year in Experiment 2; 21.69 \pm 1.08 year in Experiment 3). All were right-handed, had normal or corrected-to-normal vision, and reported no neurological history. Informed consent was obtained prior to scanning. This study was approved by a local ethics committee.

Stimuli

Stimuli consisted of digital photographs of faces with neutral or pain expressions that were taken from 16 Chinese models (8 males, aged between 19 and 24 years) and 16 Caucasian models (8 males, aged between 19 and 25 years). Based on explicit criteria of painful expressions (i.e., brow lowering, orbit tightening, and raising of the upper lip; Prkachin, 1992), one photograph of a pain expression and one of a neutral expression were selected for each model (illustrated in Fig. 1A). The luminance levels of the photographs were matched between the faces from Caucasian and Chinese models.

To validate the models' facial expressions, we asked two independent groups of participants (20 Chinese, 10 males, 21.95 \pm 1.93 year; 20 Caucasian, 10 males, 22.40 \pm 2.14 year) to evaluate the emotional intensity of each photograph along seven dimensions (i.e., fear, sadness, pain, surprise, disgust, happiness, anger) on a 7-point Likert scale (1=not at all, 7=extremely strong). They also made judgments on racial identity (Asian vs. Caucasian) of each model and rated their confidence about their judgments on a 7-point Likert scale (1=not at all confident, 7=very confident). The attractiveness of each face was also evaluated on a 7-point Likert scale (1=not at all attractive, 7=extremely attractive). One hundred and twelve photographs were presented in a random order and the order of different judgments was counterbalanced across participants. After excluding the photographs

with conflicting judgments of race, emotion, or extreme rating scores of attractiveness, 64 photographs (half with a painful expression) from 16 Caucasian and 16 Chinese models (half males in each racial group) were selected. A repeated measure analysis of variance (ANOVA) with Raters' race (Caucasian vs. Chinese), Racial group relation between raters and models (same-race vs. other-race), and Expression (Pain vs. Neutral) as within-subjects variables was conducted on rating scores of confidence for race judgments and rating scores of attractiveness (Table 1). The ANOVA did not show any significant effects ($P_s > 0.1$), suggesting comparable subjective feelings of racial identity and facial attractiveness about the Asian and Caucasian faces.

Emotion rating scores were subjected to a 2 (Raters' race: Asian vs. Caucasian) \times 2 (Racial group relation between raters and models: same-race vs. other-race) \times 2 (Expression of faces: Pain vs. Neutral) \times 7 (Emotion Dimension: fear, sadness, pain, surprise, disgust, happiness, anger) ANOVA. There were significant main effects of Expression ($F(1,38)=391.88, P < 0.001$) and Emotion Dimension ($F(6,228)=61.30, P < 0.001$). The interaction between Expression \times Emotion Dimension was significant ($F(6,228)=100.42, P < 0.001$), suggesting that the differential rating scores of pain vs. neutral expressions along the pain dimension were greater than those along other dimensions (Fig. 2). Pairwise comparisons indicated that rating scores of pain intensity of pain expressions were larger than those for other emotions ($P_s < 0.001$). However, no effect involving raters' race or the racial group relation was significant ($P_s > 0.1$). These results indicate that the emotional faces used in our study were subjectively perceived as painful rather than as portraying any of the other six emotions displayed by the Caucasians and Chinese models.

Procedure

During the electroencephalography (EEG) recordings, each photograph was presented in the center of a gray background on a 21-inch color monitor, subtending a visual angle of 3.8 \times 4.7 $^\circ$ (width \times height: 7.94 cm \times 9.92 cm) at a viewing distance of 120 cm. Each trial consisted of a face stimulus with a duration of 200 ms, which was followed by a fixation cross with a duration varying randomly between 800 ms and 1400 ms (Fig. 1B). Each experiment included 8 blocks of 128 trials (each of the 64 photographs was presented twice in a random order in each block). In Experiment 1, participants performed race judgments (Asian vs. Caucasian) on upright faces in 4 blocks and on inverted faces in 4 blocks. In Experiment 2, participants performed race judgments (Asian vs. Caucasian) in 4 blocks and pain judgments (pain vs. neutral) in 4 blocks. In Experiment 3, participants were presented with faces in blue or green T-shirts (representing fellow or opponent teams) and performed race judgments (Asian vs. Caucasian) in 8 blocks. Each block started with the presentation of an instruction that defined the task. Participants responded to each stimulus with a left or right button press using the left or right index finger. The order of stimuli (up-right vs. inverted faces in Experiment 1) and judgment tasks (race vs. pain judgments in Experiment 2) were counterbalanced across blocks. The relation between

Table 1
Subjective ratings of confidence for race judgments and attractiveness of Asian and Caucasian faces with pain and neutral expressions (mean \pm SD).

Rating (1..7)	Rater	Asian face		Caucasian face	
		Neutral	Pain	Neutral	Pain
Confidence	Asian	6.53 \pm 0.57	6.49 \pm 0.60	6.41 \pm 0.67	6.44 \pm 0.61
	Caucasian	6.78 \pm 0.37	6.60 \pm 0.63	6.73 \pm 0.47	6.70 \pm 0.36
Attractiveness	Asian	3.78 \pm 0.52	3.71 \pm 0.58	3.98 \pm 0.46	3.60 \pm 0.65
	Caucasian	3.67 \pm 0.75	3.81 \pm 0.98	3.96 \pm 0.81	3.69 \pm 0.80

FACES were rated on a 7-point Likert-scale.

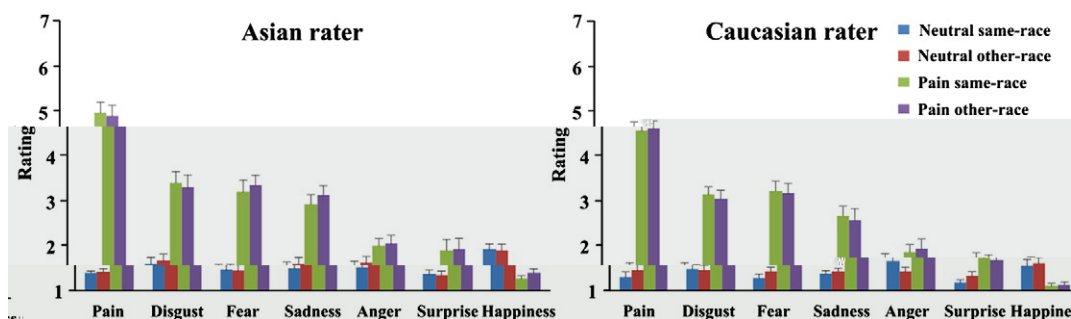


Fig. 2. Rating scores of facial expressions in seven dimensions from independent groups of Asian and Caucasian participants. Painful emotion was rated highest for photographs of pain expressions in both Asian and Caucasian participants. Error bars are standard errors.

response buttons and judgments was also counterbalanced across different blocks of trials.

After the EEG session, participants were asked to rate the intensity of the pain portrayed by each face and their own subjective feelings of unpleasantness induced by each face on a 9 point Likert scale (1 = not at all painful or unpleasant, 9 = extremely painful or unpleasant). To assess explicit subjective attitudes towards Asian and Caucasian faces, participants were asked to rate the likability of each face on a 9-point Likert scale (1 = not at all, 9 = extremely strong). Participants also completed the Interpersonal Reactivity Index (IRI) as a measure of empathy ability (Davis, 1983). The IRI is a questionnaire measure that contains four subscales including the Perspective Taking subscale that assesses the tendency to spontaneously adopt the psychological point of view of others in everyday life, the Fantasy subscale that estimates the tendency to imaginatively transpose oneself into emotional situations, the Empathic Concern subscale that assesses the tendency to experience feelings of sympathy and compassion for unfortunate others, and the Personal Distress subscale that assesses tendency to experience distress and discomfort in response to extreme distress in others (Davis, 1996, p. 57). Each subscale consists of 7 items. Participants rated each item on a 5 point scale.

Before the EEG session in Experiment 3, participants completed three learning tasks in order to remember which models were described as belonging to the same team as the participant (fellow-team members) or to a team opposing the participant's team (opponent team members). Eight Caucasian and eight Chinese models were assigned to a blue team and the other models were assigned to a green team, defined by the color of the models' T-shirts. Models' gender and the intensity of their pain expressions were matched between the two teams. Participants were informed that they had been randomly assigned to the blue or green team for a competitive game, and that they had to remember all of the fellow-team and opponent-team members. To ensure that participants believed in the existence of the game, a gender-matched experimenter was also present and assigned to the opponent team. Both the subject and the experimenter were asked to put on a blue or green t-shirt. They were then presented with neutral faces of the 32 models in colored t-shirts simultaneously and were asked to learn and remember the fellow-team and opponent-team members. To avoid the possibility that participants developed a strategy of only remembering fellow-team members, participants were told that a third team would appear in a subsequent procedure, so it was very important that they had to remember the members of both teams. This first learning task lasted for about 5 min.

In the second learning task, faces of fellow-team and opponent-team members were presented on the left and right sides of the screen, respectively (or the reverse). Participants were asked to search for a target face in a face array consisting of the fellow-team and opponent-team members by moving a frame around one of the faces. Each participant completed 4 blocks of 32 trials. In Blocks 1 and 2, the target face and the faces in the search array were matched in expression (neutral or

pain). In Blocks 3 and 4, the target face and faces in the search array were different in expression (e.g., searching for a target with a pain expression in an array of faces with neutral expressions, or the reverse). In the third learning task, each face without a colorful T-shirt was presented on a screen until the participant pressed a button to categorize the face as a fellow-team or opponent-team member. Each participant completed two blocks of categorization tasks. Each face appeared once in each block, and feedback was given after each trial. Each participant was given a memory test before and after the EEG session, respectively. The procedure of the memory test was identical to the third learning task except that only one block of trials without feedback was presented.

After the EEG session, participants were asked to complete a race version of the Implicit Association Test (IAT; Greenwald et al., 1998), where they categorized Asian faces/positive words with one key and Caucasian faces/negative words with another key in two blocks and Asian faces/negative words with one key and Caucasian faces/positive words with another key in another two blocks. Latency differences between the blocks with different response associations between faces and words reflected the relative ease of making associations between Asian versus Caucasian faces and concepts of good and bad. According to the established algorithm of the latencies (Greenwald et al., 2003), a positive IAT D score indicates that compared to Caucasians, Asians are associated with good rather than bad while a negative IAT D score indicates that, compared to Caucasians, Asians are associated with bad rather than good.

ERP data recording and analysis

The EEG was continuously recorded from 62 scalp electrodes that were mounted on an elastic cap in accordance with the extended 10.20 system and were referenced to the average of the left and right mastoid electrodes. The electrode impedance was kept less than 5 k Ω . Eye blinks and vertical eye movements were monitored with 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 was amplified (band pass 0.1–100 Hz) and digitized at a sampling rate of 250 Hz. The ERPs in each condition were averaged separately off-line with an epoch beginning 200 ms before stimulus onset and continuing for 1200 ms. Trials contaminated by eye blinks, eye movements, muscle potentials exceeding ± 50 μ V at any electrode, or response errors were excluded from the average. This resulted in rejection of 16.1% trials in Experiment 1, 15.4% in Experiment 2, and 18.1% in Experiment 3 (see the number of trials accepted for average in each condition in Supplementary Table S1). The baseline for ERP measurements was the mean voltage of a 200 ms pre-stimulus interval and the latency was measured relative to the stimulus onset. Mean amplitudes of each ERP component were calculated at electrodes selected from frontal (Fz, FCz, F3, F4, FC3, FC4), central (Cz, C3, C4), parietal (Pz, P4), and occipito-temporal (PO7, PO8, P7, P8) regions. Behavioral

performances and ERPs were subjected to ANOVAs with Expression (pain vs. neutral) and models' Race (Asian vs. Caucasian) as within-subjects variables in Experiments 1 and 2. Another within-subjects variable was also included in Experiment 1 (Orientation (upright vs. inverted faces)), Experiment 2 (Task (race vs. pain judgments)), and Experiment 3 (Team (fellow vs. opponent team members)), respectively.

Both voltage topography and the standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) (Pascual-Marqui, 2002) were used to estimate potential sources of empathic neural responses. sLORETA is a linear method of computing statistical maps from EEG data that reveal locations of the underlying source processes and do not require a priori hypotheses regarding the field distribution of the active sources. We performed the analysis using sLORETA to assess the 3D current source of neural activity that differentiated between ERPs to pain and neutral expressions. A boundary element model was first created with about 5000 nodes from a realistic head model. Statistical nonparametric mapping was calculated in a specific time window to estimate the source that differentiated ERPs to pain and neutral expressions. The log of the F ratio of averages was used and considered with a 0.95 level of significance.

Results

Experiment 1: racial biases in empathic neural responses to pain expressions

Experiment 1 recorded ERPs during race judgments (Asian vs. Caucasian) on upright and inverted faces with pain or neutral expressions in separate blocks of trials. Behavioral performances are shown in Table 2. Response accuracies were high (>88%). ANOVAs of reaction times (RTs) showed significant main effects of Race ($F(1,15)=10.68, P<0.01$), Expression ($F(1,15)=11.99, P<0.005$), and Task ($F(1,15)=9.41, P<0.01$). There was also a significant three-way interaction of Race×Expression×Orientation ($F(1,15)=7.60, P<0.05$). For upright faces there was a significant interaction of Race×Expression on RTs to upright faces ($F(1,15)=6.23, P=0.024$). Post hoc analyses showed that RTs were longer to pain than to neutral expressions in Asian faces ($F(1,15)=12.57, P<0.005$) but did not differ between pain and neutral expressions in Caucasian faces ($F<1$). ANOVAs of RTs to inverted faces showed only a significant main effect of Race ($F(1,15)=8.623, P<0.01$) as subjects responded faster to Caucasian than to Asian faces. Neither the main effect of Expressions nor its interaction with Race was significant ($F(1,15)=3.48$ and $0.95, P_s>0.05$).

Rating scores of pain intensity and self-unpleasantness were higher for pain than for neutral expressions ($F(1,15)=297.06$ and $54.6, P_s<0.001$), however, the effects of facial expression on rating scores did not differ between Asian and Caucasian faces ($F(1,15)=0.06$ and $0.63, P_s>0.1$, Table 3). Likability ratings were higher for neutral than pain expressions ($F(1,15)=8.58, P<0.01$) but did not differ significantly between Asian and Caucasian faces ($F(1,15)=1.204, P>0.1$). The IAT D index did not significantly differ from zero (mean±SD=0.02±0.28, $t(15)=0.29, P>0.5$), suggesting comparable implicit attitudes toward Asians and Caucasians.

The ERPs to faces were characterized by a negative wave at 84..116 ms (N1) and a positive deflection at 128..188 ms (P2) over the frontal..central area. These were followed by a negative wave at 200..300 ms (N2) over the frontal region and a long-latency positivity at 400..700 ms (P3) over the central/parietal area. Face stimuli also elicited a posterior positivity at 88..148 ms (P1) and a negativity at 140..180 ms (N170). Similar to the previous fMRI (e.g., Singer et al., 2004) and ERP (e.g., Fan and Han, 2008) studies, empathic neural responses were defined by increased neural activity to painful compared to neutral stimuli. Fig. 3A illustrates grand-averaged ERPs to upright faces.

As ANOVAs of the P2 amplitude at 128..188 ms showed significant interactions of Race×Expression×Orientation ($F(1,15)=4.95..5.71, P_s<0.05$)¹, we then analyzed the P2 amplitude separately for upright and inverted faces. For upright faces the P2 amplitude was greater to pain than neutral expressions ($F(1,15)=31.03..37.53, P_s<0.001$) and greater to Caucasian than Asian faces ($F(1,15)=19.56..29.52, P_s<0.001$, see Supplementary Table S2 for P2 amplitudes). The sLORETA analysis suggested that the neural activity in the P2 time window that differentiated between pain and neutral expressions of Asian faces had potential sources in the dorsal ACC and supplementary motor cortex (peak MNI coordinates: $-5, 40, 25$, Fig. 3B). Importantly, there were significant interactions of Race×Expression on the P2 amplitude to upright faces ($F(1,15)=14.48..26.27, P_s<0.005$). Post hoc analyses further confirmed that the P2 amplitude was larger to pain than to neutral expressions of Asian faces ($F(1,15)=42.22..72.24, P_s<0.001$) but not of Caucasian faces ($F(1,15)=0.10..0.79, P_s>0.05$, Figs. 3A and D). Inverted Caucasian faces elicited larger P2 relative to inverted Asian faces ($F(1,15)=11.25..18.46, P_s<0.005$, Fig. 3C). However, neither the main effect of Expression ($F(1,15)=0.41..4.12, P_s>0.05$) nor its interaction with Race ($F(1,15)=0.06..1.54, P_s>0.05$) on the P2 amplitude to inverted faces was significant, suggesting that differences in low level visual features between Asian and Caucasian faces, which were

Table 3
Rating scores (mean±SD) of upright faces in Experiment 1.

	Pain intensity		Unpleasantness		Likability	
	Neutral	Pain	Neutral	Pain	Neutral	Pain
Asian face	1.92±1.17	6.4±1.14	2.45±1.35	4.88±1.84	4.62±0.93	4.05±0.97
Caucasian face	1.85±1.05	6.5±1.04	2.31±1.2	4.88±1.58	4.80±0.92	4.07±0.85

scores of pain intensity and self-unpleasantness and the IRI scores. The empathic neural responses in the P2 time window were positively correlated with subjective ratings of self-unpleasantness ($r=0.506-.0627, P<0.05$) and subjective ratings of the empathic concern subscale ($r=0.543-.0624, P_s<0.05$, Fig. 5). However, the P2 RBE effect did not correlate with individual's explicit or implicit attitude biases ($r= -0.315-.0031, P_s>0.1$).

Experiment 2: cognitive strategy and racial bias in empathy

Experiment 2 recorded ERPs during race judgments (Asian vs. Caucasian) and pain judgments (pain vs. neutral expression) in separate blocks of trials on Asian and Caucasian faces with pain and neutral expressions. Behavioral performances are shown in Table 4. Response accuracies were high (>91%). ANOVAs of RTs showed significant main effect of Expression ($F(1,15)=11.60, P <0.005$) and significant interaction of Task and Expression ($F(1,15)=29.66, P <0.001$), suggesting that participants responded faster to neutral than to pain expressions during race judgments whereas a reverse pattern was true during pain judgments.

ANOVAs of the P2 amplitude at 128–188 ms showed significant three-way interactions of Race×Expression×Task ($F(1,15)=5.58 \dots 7.34, P<0.05$, see Supplementary Table S4 for P2 amplitudes). We then analyzed the P2 amplitudes in the race and pain judgment tasks separately. Similar to Experiment 1, race judgments of pain vs. neutral expressions elicited larger P2 amplitudes at 128–188 ms over the frontal/central area ($F(1,15)=4.64 \dots 7.28, P_s<0.05$, Fig. 6A).

Relative to Asian faces, Caucasian faces elicited larger P2 amplitudes ($F(1,15)=20.59 \dots 33.70, P_s<0.001$). ANOVAs of the P2 amplitude to race judgments also showed significant interactions of Race×Expression ($F(1,15)=9.40 \dots 21.14, P_s<0.01$). Post hoc analyses confirmed that the P2 amplitude was enlarged to pain compared to neutral expressions in Asian faces ($F(1,15)=7.75 \dots 20.09, P_s<0.05$) but not in Caucasian faces ($P_s>0.05$, Fig. 6E). Pain judgments of pain vs. neutral expressions elicited a larger P2 amplitude at 128–188 ms over the frontal/central electrodes ($F(1,15)=20.29 \dots 28.53, P_s<0.001$, Fig. 6C). Relative to Asian faces, Caucasian faces elicited a larger P2 amplitude ($F(1,15)=24.48 \dots 37.62, P_s<0.001$). However, there was no significant interaction of Race×Expression on the P2 amplitudes during pain judgments ($F(1,15)=0.04 \dots 2.68, P_s>0.1$, Fig. 6E), suggesting comparable neural responses to pain vs. neutral expressions of Asian and Caucasian faces. The sLORETA analysis suggested that the neural activity in the P2 time window that differentiated between pain and neutral expressions of Asian and Caucasian faces in both the race and pain judgment tasks had potential sources in the dorsal ACC and supplementary motor cortex (peak MNI coordinates for race judgments: $-10, 40, 20$; Fig. 6B; peak MNI coordinates for pain judgments: $0, 25, 20$; Fig. 6D).

ANOVAs of the N2 amplitudes at 200–300 ms showed a significant effect of Expression ($F(1,15)=23.60 \dots 36.41, P_s<0.001$), as pain vs. neutral expressions induced a positive shift of the N2 amplitudes in this time window over the frontal/central area (Fig. 6A, see Supplementary Table S5 for N2 amplitudes). However, the three-way interaction of Race×Expression×Task did not reach significance ($F(1,15)=$

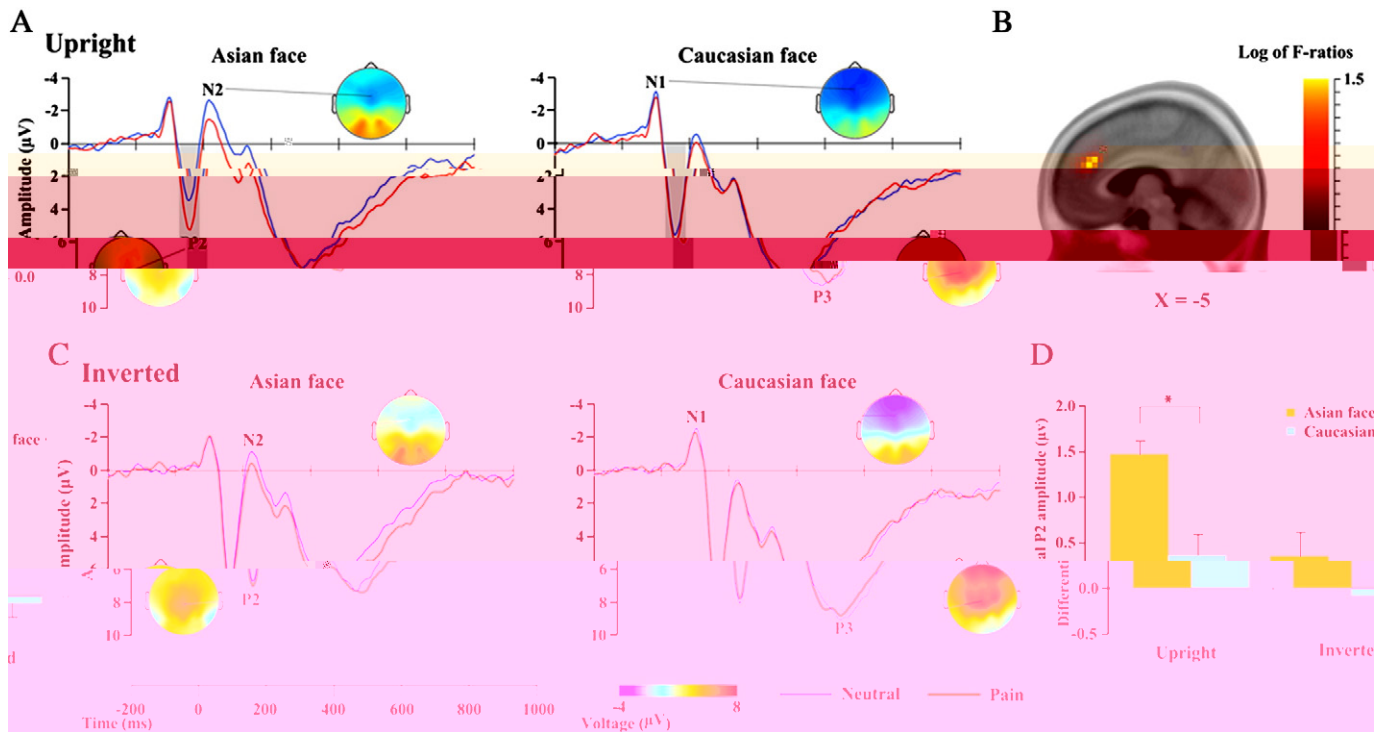


Fig. 3. ERP results in Experiment 1. (A) Grand-averaged ERPs to upright faces recorded at the electrode FCz. Voltage topographies illustrate the scalp distribution of the maximum amplitude of each ERP component. (B) Source estimation of the neural activity that differentiated between pain and neutral expressions of Asian faces at 148 ms. (C) Grand-averaged ERPs to inverted faces at FCz. (D) The differential P2 amplitudes at 128–188 ms to pain vs. neutral expressions at FCz. Error bars are standard errors. The gray rectangles in (A) and (C) indicate the time windows where the P2 amplitudes were measured.

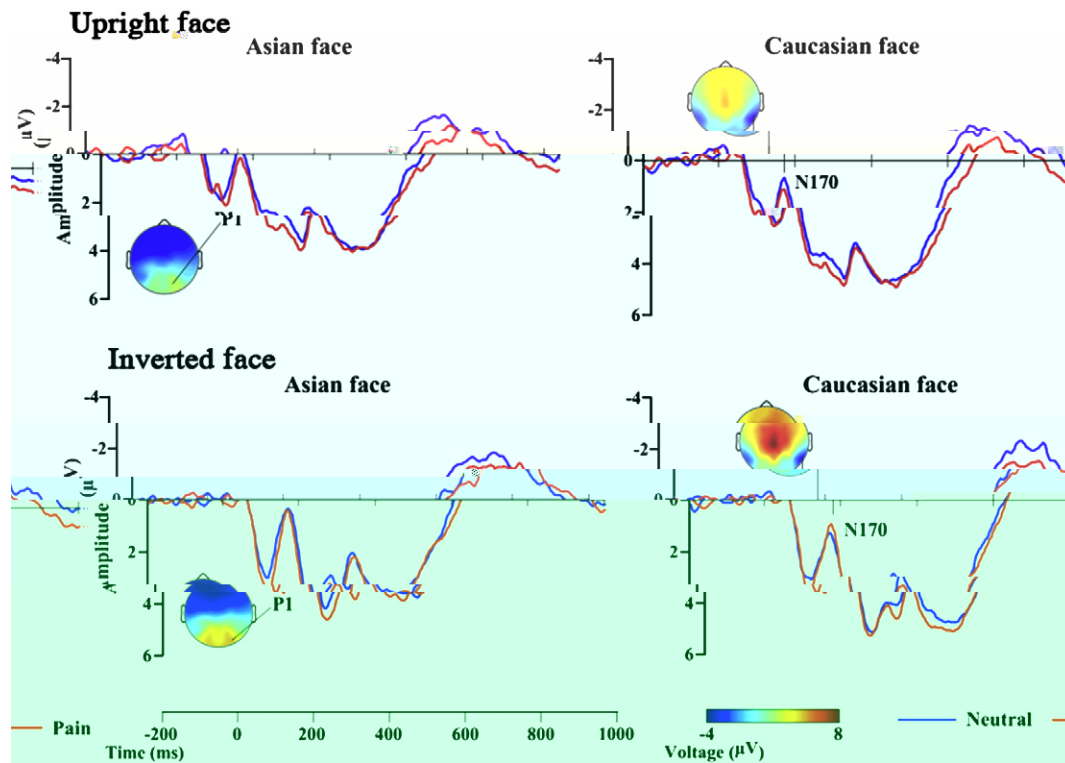


Fig. 4. ERPs recorded at the posterior electrode in Experiment 1. Grand-averaged ERPs to upright and inverted faces recorded at the electrode P8 in Experiment 1.

1.28-.337, $P_s > 0.05$). ANOVAS of the P3 amplitude only showed a significant main effect of Race ($F(1,15) = 6.26$, $P_s < 0.05$). Relative to Asian faces, Caucasian faces elicited a positive shift of the ERPs at 400-.700 ms.

Because actively contemplating others' psychological experiences through perspective taking attenuates automatic expressions of racial bias (Todd et al., 2011), we tested if attention manipulation was more efficient at reducing the RBE in those with better perspective-taking abilities. Indeed, we found that the increased neural responses to pain vs. neutral expressions of Caucasian faces as a function of task (i.e., P2 amplitude to pain vs. neutral expressions during pain judgments minus P2 amplitude to pain vs. neutral expressions during race judgments) were positively correlated with ratings on the perspective-taking subscale of the IRI ($r = 0.543$ -.0666, $P_s < 0.05$; Fig. 6F). This

suggests that the pain judgment task increased neural responses to pain vs. neutral expressions of Caucasian faces to a greater degree in those with better perspective-taking ability. Similar analysis of the neural responses to pain vs. neutral expressions of Asian faces as a function of task did not produce such effect ($r = -0.156$ -.0.129, $P_s > 0.1$).

Experiment 3: Intergroup relationship and racial bias in empathy

Experiment 3 investigated if including other-race members in one's own (fellow) team for an oppositional game can increase empathic neural responses to their faces. After the learning tasks and the EEG recording session, participants had to recognize fellow and opponent team members without T-shirts and showed high accuracy at this task (>90%). Behavioral performances are shown in Table 5.

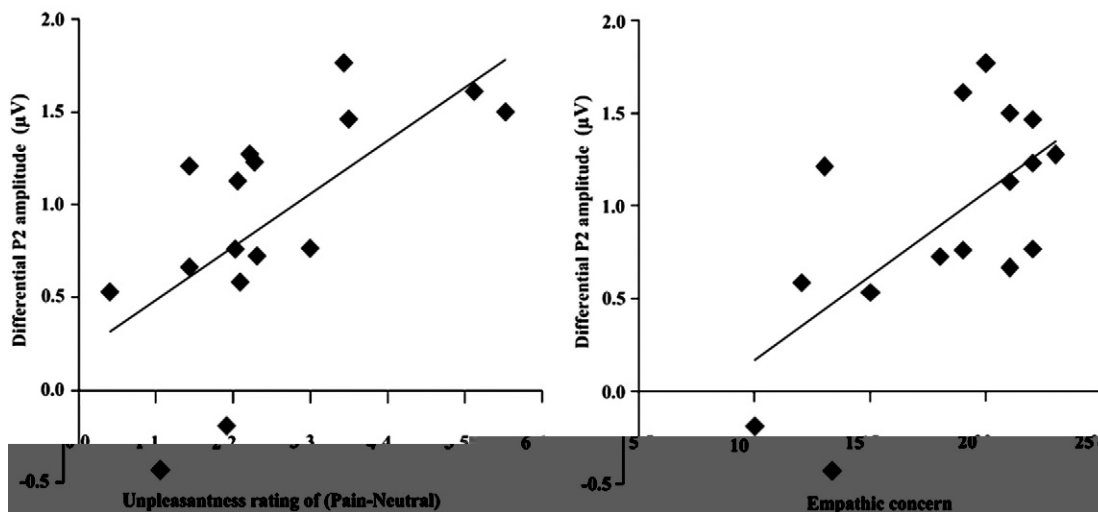


Fig. 5. Correlation between subjective ratings of self-unpleasantness and differential P2 amplitudes to pain vs. neutral expressions and correlation between subjective ratings of empathic concern and differential P2 amplitudes to pain vs. neutral expressions at FCz.

Table 4
Behavioral performances (mean±SD) in Experiment 2.

		Race judgment		Pain judgment	
		Neutral	Pain	Neutral	Pain
Reaction time (ms)	Asian face	515±49	528±47	512±54	499±52
	Caucasian face	508±47	506±48	518±57	494±57
Accuracy (%)	Asian face	93.1±5.55	91.2±5.76	92.2±5.57	91.3±5.12
	Caucasian face	92.3±5.04	92.6±4.89	91.1±5.75	91.6±4.29

Response accuracies of race judgments were high during EEG recordings (>93%). ANOVAs of RT showed a significant main effect of Expression ($F(1,15)=17.24, P < 0.001$) and significant interaction of Race and Expression ($F(1,15)=4.91, P < 0.05$), suggesting that pain expression slowed response speed and this effect was greater on

Asian than on Caucasian faces. Subjective ratings of likability were greater to fellow compared to opponent team faces ($F(1,15)=10.343, P < 0.01$), but did not differ between Asians and Caucasians ($F(1,15)=0.01, P > 0.1, Table 6$).

ANOVAs of the P2 amplitude at 128–188 ms showed significant three-way interactions of Race×Expression×Team ($F(1,15)=5.72, P < 0.05$). We then analyzed the P2 amplitudes to the fellow-team and opponent-team faces, respectively. Pain expressions of opponent team faces elicited larger P2 amplitudes at 128–188 ms over the frontal/central area compared to neutral expressions ($F(1,15)=5.25, P < 0.05, Fig. 7A$, see Supplementary Table S6 for P2 amplitudes). Caucasian faces elicited a larger P2 amplitude relative to Asian faces ($F(1,15)=54.29, P < 0.001$). There were significant interactions of Race×Expression on the P2 amplitude ($F(1,15)=5.16, P < 0.05$). Post hoc analyses confirmed larger P2 amplitudes to pain than neutral expressions in Asian faces ($F(1,15)=10.51, P < 0.005$) but not in Caucasian faces ($F(1,15)=0.06, P > 0.1$).

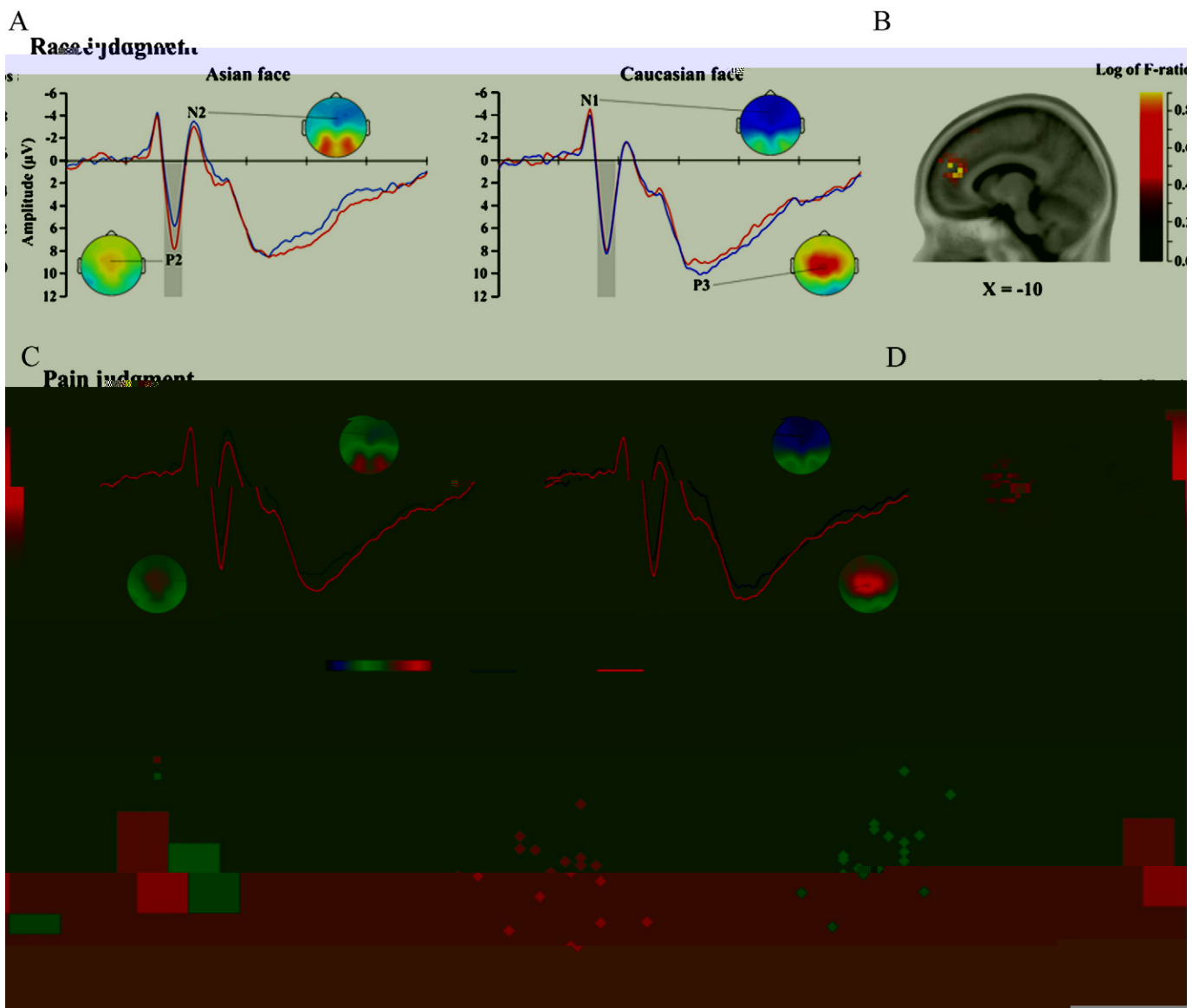


Fig. 6. ERP results in Experiment 2. (A) Grand-averaged ERPs during race judgments recorded at FCz. (B) Source estimation of the neural activity that differentiated between pain and neutral expressions of Asian faces during race judgments at 140 ms. (C) Grand-averaged ERPs during pain judgments recorded at FCz. (D) Source estimation of the neural activity that differentiated between pain and neutral expressions during pain judgments at 136 ms. (E) The differential P2 amplitudes at 128–188 ms to pain vs. neutral expressions at FCz during race and pain judgments. Error bars are standard errors. (F) The correlation between ratings of perspective-taking subscale and the differential empathic responses in the P2 time window at FCz (defined by (pain-neutral) pain judgments minus (pain-neutral) race judgments). The gray rectangles in (A) and (C) indicate the time windows where the P2 amplitudes were measured.

Table 5
Behavioral performances (mean±SD) in Experiment 3.

		Fellow team		Opponent team	
		Neutral	Pain	Neutral	Pain
Reaction time (ms)	Asian face	514±59	524±58	515±54	526±61
	Caucasian face	513±59	520±55	513±50	512±53
Accuracy (%)	Asian face	95.3±4.02	94.8±3.77	95.0±4.76	94.1±4.16
	Caucasian face	93.9±3.45	94.6±4.29	93.7±3.34	93.8±3.87

Fig. 7E). Relative to neutral expressions, pain expressions of fellow team faces elicited a larger P2 amplitude at the frontal/central electrodes ($F(1,15)=18.48$, $p=.031$, $P_s<0.001$, Fig. 7C). The P2 amplitude was larger to Caucasian than to Asian faces ($F(1,15)=28.77$, $p=.035$, $P_s<0.001$). However, the interaction of Race×Expression was not significant ($F(1,15)=0.04$, $p=.268$, $P_s>0.1$, Fig. 7E), suggesting comparable empathic neural responses to Asian and Caucasian faces of the fellow team members. Separate analyses of the P2 amplitude also showed a significant interaction of Expression×Team for Caucasian faces ($F(1,15)=8.86$, $p=.013$, $P_s<0.01$) but not to Asian faces ($F(1,15)=0.09$, $p=.055$, $P_s>0.05$). Thus the manipulation of intergroup relationships increased empathic neural responses to other-race faces but did not affect empathic neural responses to same-race faces. The sLORETA analysis suggested that the neural activity in the P2 time window that differentiated between pain and neutral expressions of Asian faces from the opponent team (peak MNI coordinates: 5, 35, 25; Fig. 7B) or of both Asian and Caucasian faces from the fellow team (peak MNI coordinates: -10, 40, 30; Fig. 7D) had potential sources in the dorsal ACC and supplementary motor cortex.

ANOVAs of the N2 amplitude also showed a significant three-way interaction of Race×Expression×Team at 200–300 ms ($F(1,15)=5.38$, $p=.013$, $P_s<0.05$). Separate analyses revealed that pain expressions of opponent team faces elicited a positive shift of the N2 amplitudes at 200–300 ms over the frontal/central area compared to neutral expressions ($F(1,15)=4.60$, $p=.050$, $P_s<0.05$, Fig. 7A, see Supplementary Table S7 for N2 amplitudes). Moreover, there was a significant interaction of Race×Expression ($F(1,15)=5.03$, $p=.059$, $P_s<0.05$) as the positive shift of the N2 amplitude was evident for Asian faces ($F(1,15)=5.81$, $p=.013$, $P_s<0.05$) but not for Caucasian faces ($F(1,15)=0.01$, $p=.121$, $P_s>0.1$). Relative to neutral expressions, pain expressions of fellow team faces elicited a positive shift of the N2 ($F(1,15)=4.73$, $p=.033$, $P_s<0.05$, Fig. 7C). However, this effect did not differ between Asian and Caucasian faces ($F(1,15)=0.96$, $p=.305$, $P_s>0.1$). Separate analyses of the N2 amplitude confirmed a significant interaction of Expression×Team for Caucasian faces ($F(1,15)=5.04$, $p=.067$, $P_s<0.05$) but not for Asian faces ($F(1,15)=0.97$, $p=.256$, $P_s>0.1$). Similarly, the manipulation of intergroup relationships increased empathic neural responses to other-race faces but did not affect empathic neural responses to same-race faces in the N2 time window. ANOVAs of the P3 amplitude at 400–700 ms only showed a significant main effect of Race ($F(1,15)=25.37$, $p=.036$, $P_s<0.001$). Caucasian faces elicited a positive shift of the ERPs at 400–700 ms compared to Asian faces.

To test whether participants who were slowed during racial face categorization (implicating enhanced individuation of faces) by intergroup relationship manipulations showed greater improvement of

empathy for Caucasian faces, we calculated the correlation between differential RTs to fellow vs. opponent team faces and the differential empathic responses in the P2 time window to fellow vs. opponent team faces. This resulted in a significant correlation for Caucasian faces ($r=0.511$, $p=.032$, $P_s<0.05$) but not for Asian faces ($r=-0.090$, $p=.312$, $P_s>0.1$, Fig. 7F).

Discussion

Experiment 1 first identified empathic neural responses in ERPs to pain expression. Pain expressions slowed RTs, increased P2 amplitudes, and induced a positive shift of the N2 amplitudes, similar to the previous findings (Decety et al., 2010; Fan and Han, 2008). Moreover, Experiment 1 showed ERP evidence for the RBE, i.e., the effects of pain expressions on RTs, P2 and N2 amplitudes were significantly reduced for other-race than same-race faces. Pain expressions of other-race faces failed to modulate either RTs or ERP amplitudes. Slowed RTs to pain compared to neutral expressions suggest that the pain expression distracted participants' attention to race and interfered with their responses to race judgments. However, this effect on RTs was stronger for Asian than for Caucasian faces, suggesting that participants were more sensitive to pain expression of Asian faces than that of Caucasian faces. The source analysis suggested that the P2 modulation by facial expressions might arise from the ACC and supplementary motor cortex, consistent with previous fMRI findings (Gu and Han, 2007; Han et al., 2009; Jackson et al., 2005; Saarela et al., 2007; Singer et al., 2004). Moreover, the P2 effect was correlated with subjective feelings of unpleasantness induced by perceived pain and with individuals' empathy traits, suggesting a key role of the neural activity in the P2 time window in mediating the understanding and sharing of others' painful feelings.

Consistent with the previous fMRI findings (Xu et al., 2009), our ERP results indicated increased empathic neural responses to pain expressions of racial in-group compared to racial out-group members. This effect was observed mainly in the P2 time window. Interestingly, RBE-related brain activity occurred when participants reported similar subjective feelings of perceived pain intensity and self-unpleasantness associated with same-race and other-race faces, similar to the results reported in our previous research (Xu et al., 2009). Thus RBE-related brain activity may occur without conscious awareness. In addition, the RBE-related brain activity cannot be simply attributed to differences in attitudes toward same-race and other-race members because neither the explicit nor implicit attitude measurements showed any racial in-group bias in our participants. Categorical processing of social targets may automatically symbolize other-race members and individuate same-race faces (Valentine and Endo, 1992; Vizioli et al., 2010), resulting in different neurocognitive processing of emotional states of racial in-group compared to out-group members.

Most importantly, we demonstrated that RBE-related brain activity was reduced by manipulations of cognitive strategy and intergroup relationships. Experiment 2 showed that RBE-related brain activity was eliminated when participants paid attention to the observed individual's feelings of pain. Moreover, the RBE was reduced by increasing empathic neural responses to other-race faces and this effect was predicted by self-reported perspective-taking abilities. Our ERP findings seemed to be inconsistent with the previous fMRI research that observed RBE-related ACC activity in a similar pain judgment task (Mathur et al., 2010; Xu et al., 2009). However, the fMRI studies used video clips with long durations of either painful stimulation applied to neutral faces (3-s video clips) (Xu et al., 2009) or complicated scenes (2.5-s pictures) (Mathur et al., 2010) and required pain judgments after the offset of the stimuli. Thus categorization of faces based on race, which occurred around 150 ms after stimulus onset in our study, might have occurred prior to pain judgments in the fMRI studies. The RBE-related P2 effect observed here may reflect modulations of the early automatic empathic neural response (Fan and Han,

Table 6
Likability ratings (mean±SD) in Experiment 3.

	Fellow team		Opponent	
	Asian	Caucasian	Asian	Caucasian
Neutral expression	5.83±0.72	5.94±0.88	5.31±0.60	5.35±0.79
Pain expression	5.96±1.05	5.78±1.07	5.42±0.99	5.29±0.97

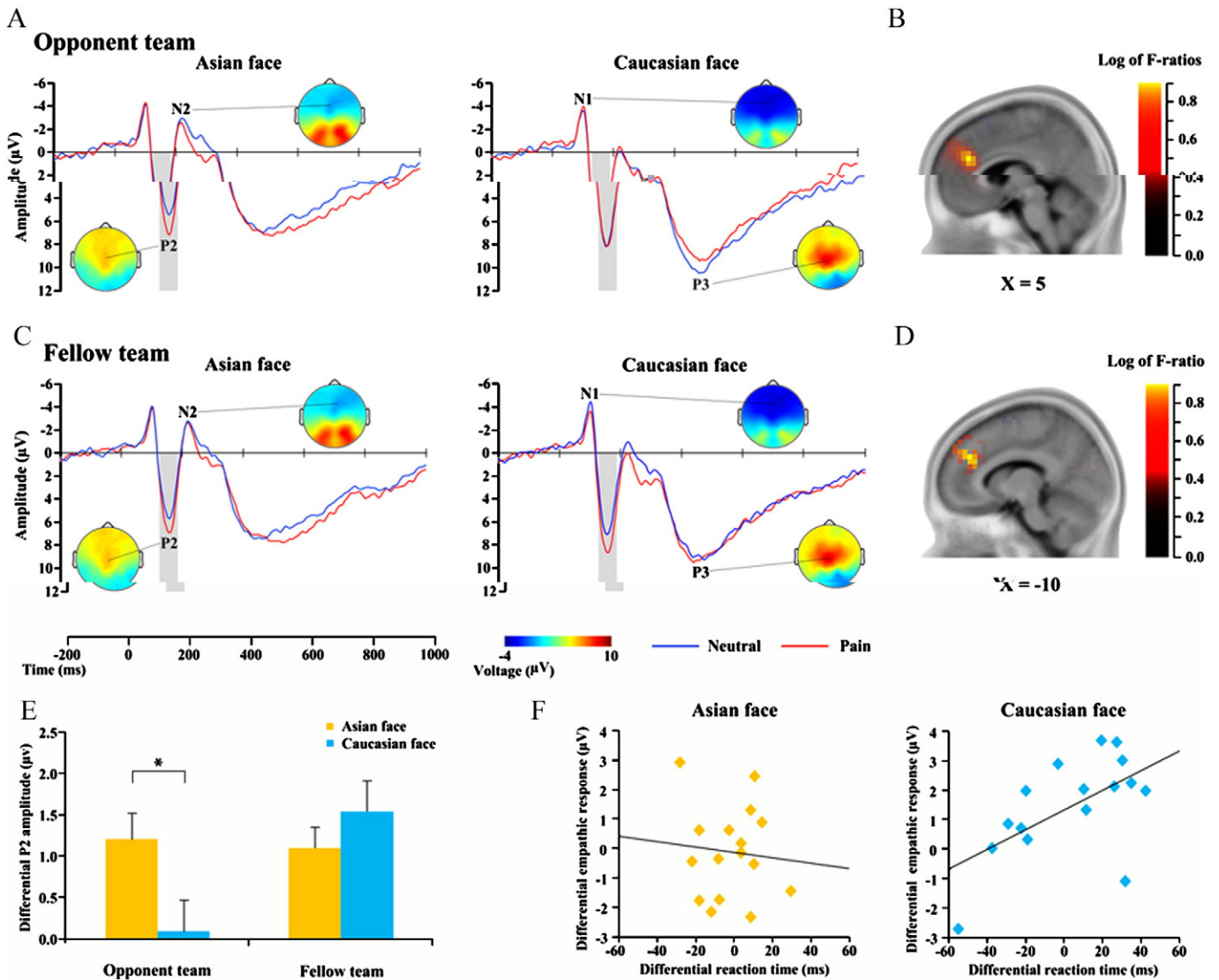


Fig. 7. ERP results in Experiment 3. (A) Grand-averaged ERPs to opponent team faces recorded at FCz. (B) Source estimation of the neural activity that differentiated between pain and neutral expressions of Asian faces from the opponent team faces at 156 ms. (C) Grand-averaged ERPs to fellow team faces recorded at FCz. (D) Source estimation of the neural activity that differentiated between pain and neutral expressions of fellow team faces at 148 ms. (E) The differential P2 amplitudes at 128–188 ms to pain vs. neutral expressions at FCz. Error bars are standard errors. (F) The correlation between differential RTs (mean RTs to fellow team faces minus mean RTs to opponent team faces) and the differential empathic responses in the P2 time window at FCz ((pain-neutral)_{fellow team} minus (pain-neutral)_{opponent team}). The gray rectangles in (A) and (C) indicate the time windows where the P2 amplitudes were measured.

2008). It is well known that pain perception is modulated by attention such that focusing on pain enhances pain perception (see [Villemure and Bushnell, 2002](#) for a review). Attention-related modulations of nociceptive responses have been observed in both sensory and limbic cortical areas, including SI, SII, ACC and insular cortices ([Bushnell et al., 1999](#); [Petrovic et al., 2000](#); [Peyron et al., 1999](#)). However, the effect of attention on empathic neural responses to pain expression remains unclear. Several fMRI studies similarly observed activation in the affect nodes of the pain matrix (e.g., the ACC) regardless of task demand. Attentively viewing video clips ([Botvinick et al., 2005](#); [Lamm et al., 2007](#)) or photographs ([Saarela et al., 2007](#)) that showed pain vs. neutral expressions increased the activity in the ACC and bilateral insula. Viewing video clips of pain expression similarly activated the ACC even though participants' attention was distracted from pain expression by being asked to perform gender discrimination of painful faces ([Simon et al., 2006](#)). Our ERP results suggest that, similar to first-person pain perception, empathy for others' pain expression can be enhanced by attention in a social context where racial in-group and out-group

members are present simultaneously and such effect occurs mainly to the racial out-group members.

Experiment 3 showed that changing the intergroup relationships between observers and targets by enclosing other-race models into one's own team eliminated RBE-related brain activity due to increases of empathic neural responses to other-race faces. By assigning participants to novel groups and providing equal exposure to fellow and opponent team faces, Experiment 3 minimized the roles of familiarity and novelty as causal variables in the modulation of RBE-related neural activity. Moreover, decreasing RBE-related brain activity via manipulations of intergroup relationships did not require explicit attention to team membership since race judgments rather than team membership judgments were conducted in Experiment 3.

Together, our findings support the proposal that manipulations of cognitive strategies and intergroup relationships may increase references to an individual's personal situation when perceiving painful expressions in other-race faces and, consequently, lead to increased empathic neural responses to other-race individuals. Our findings

demonstrate that the RBE is not inevitable and that manipulations of cognitive strategies and intergroup relationships can reduce RBE-related brain activity effectively.

The results of Experiment 3 indicate that, while the novel in-group favoritism increased empathic neural responses to other-race faces, the novel out-group derogation did not reduce empathic responses to same-race faces as the P2 effect to same-race faces was comparable for the fellow and opponent teams. Prior fMRI research also showed that fusiform and amygdala activity increased when observing neutral faces of one's own mixed-race (black and white) team compared to observing neutral faces of a mixed-race opponent team in white participants (Van Bavel et al., 2008), suggesting enhanced processing of in-group members due to greater individuation (Rhodes et al., 2004) of in-group relative to out-group members. It appears that the temporary group relationships built through team assignment can overcome the effect of racial-group relationships on the fusiform and amygdala activity to other-race faces (Golby et al., 2001; Lieberman et al., 2005). Our ERP results suggest that the temporary group manipulation did not influence empathic neural responses to same-race faces. Thus the affective processing of same-race faces, which may occur early and automatically during face processing (Ito and Bartholow, 2009), cannot be deteriorated by simply assigning same-race faces to an opponent team for competition.

Interestingly, we found that the RBE was eliminated even when the P2 amplitudes differentiated between same-race and other-race faces. Previous studies found larger N1 and P2 amplitudes to black than white faces in white participants (Ito and Bartholow, 2009; Kubota and Ito, 2007) whereas the current work showed a larger P2 amplitude to Caucasian than Asian faces in Chinese participants. The race effect observed in ERPs in our study occurred later than that observed in Kubota and Ito (2007) possibly due to the difference in either the stimuli or participants, or both. However, our results indicate that the neural distinction between same-race and other-race faces does not necessarily result in RBE-related brain activity to painful expressions. Thus RBE-related neural processes may be dissociated from those involved in the recognition of racial faces. Consistent with this, perception of same-race vs. other-race faces increased fusiform and amygdala activity (Golby et al., 2001; Lieberman et al., 2005), whereas the RBE was manifested in the modulation of ACC and sensorimotor activity (Avenanti et al., 2010; Xu et al., 2009).

There may be alternative accounts of RBE-related brain activity. For example, given the in-group familiarity or advantage in recognizing emotions expressed by same-race members (Elfenbein and Ambady, 2002

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2012.04.028> .

References

- Avenanti, A., Sirigu, A., Aglioti, S.M., 2010. Racial bias reduces empathic sensorimotor resonance with other-race pain. *Curr. Biol.* 20, 1018–1022.
- Batson, C.D., 2009. These things called empathy: Eight related but distinct phenomena. In: Decety, Jean, Ickes, William (Eds.), *The social neuroscience of empathy*. The MIT Press, Cambridge, Massachusetts, pp. 3–15.
- Batson, C.D., Ahmad, N.Y., 2009. Using empathy to improve intergroup attitudes and relations. *Soc. Issues Policy Rev.* 3, 141–177.
- Botvinick, M., Jha, A.P., Bylsma, L.M., Fabian, S.A., Solomon, P.E., Prkachin, K.M., 2005. Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage* 25, 312–319.
- Bushnell, M.C., Duncan, G.H., Hofbauer, R.K., Ha, B., Chen, J., Carrier, B., 1999. Pain perception: is there a role for primary somatosensory cortex? *Proc. Natl. Acad. Sci. U. S. A.* 96, 7705–7709.
- Cheon, B.K., Im, D.M., Harada, T., Kim, J.S., Mathur, V.A., Scimeca, J.M., Parrish, T.B., Park, H.W., Chiao, J.Y., 2011. Cultural influences on neural basis of intergroup empathy. *Neuroimage* 57, 642–650.