Are women more empathetic than men? Questionnaire and EEG estimations of sex/gender differences in empathic ability

Chenyu Pang, Wenxin Li, Yuqing Zhou, Tianyu Gao, and Shihui Han [ID]

School of Psychological and Cognitive Sciences, PKU-IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100080, China
Correspondence should be addressed to Shihui Han, School of Psychological and Cognitive Sciences, Peking University, 52 Haidian Road, Beijing 100080, China.
E-mail: shan@pku.edu.cn.
Chenyu Pang, Wenxin Li, Yuqing Zhou and Tianyu Gao contributed equally to this work.

Abstract

The debate regarding whether women are more empathetic than men has broad scientific, social and clinical implications. However, previous independent questionnaires and brain imaging studies that tested different samples reported inconsistent results regarding sex/gender differences in empathic ability. We conducted three studies to investigate sex/gender differences in empathic ability using large-sample questionnaires and electroencephalography (EEG) measures. We showed that the estimation of empathic ability using the Interpersonal Reactivity Index questionnaire showed higher rating scores in women than in men in all studies. However, our EEG measures of empathy, indexed by both phase-locked and non–phased-locked neural responses to others' painful (*vs* neutral) facial expressions, support a null hypothesis of the sex/gender difference in empathic ability. In addition, we showed evidence that priming social expectations of women and men's ability to share and care about others' feelings eliminated the sex/gender difference in questionnaire measures of empathic ability. Our large-sample EEG results challenge the notion of women's superiority in empathy that is built based on subjective questionnaire measures that are sensitive to social desirability. Our findings indicate that whether the notion of women's superiority in empathic ability reflects a biological/social difference between women and men or a gender-role stereotype remains an open question.

Keywords: electroencephalography; empathy; questionnaire; sex/gender; social expectation

Introduction

Empathy refers to the ability to understand and share others' emotional states and plays a fundamental role in social behavior (Eisenberg and Fabes, 1990; Decety and Jackson, 2004; Hoffman, 2008). A widely distributed notion related to empathy is that women are more empathetic than men. Does this notion reflect the fact that women and men recruit distinct neurocognitive processes of perceived emotional states of others or a gender-role stereotype that is exaggerated by subjective estimation of empathy? To clarify this issue is critical for understanding the impact of biological sex on a psychological ability that is highly related to social behaviors and has broad social and clinical implications (Batson, 1991, 2011; Baron-Cohen *et al.*, 2005). Studies of sex/gender differences in empathic ability that employed different approaches have revealed inconsistent results during the last four decades (Eisenberg and Lennon, 1983; Christov-Moore *et al.*, 2014; Murphy and Lilienfeld, 2019), urging a deeper understanding of the underpinnings of previous contradictory findings.

Because 'sex' and 'gender', respectively, refer to the inflexible biological component and the psychosocial manifestation of human male–female differences (National Institutes of Health, 2015) and previous studies of empathy commonly assigned participants to female and male groups based on self-identities, we adopt the term 'sex/gender' to label populations of male and female participants in the current study, similar to previous research (Eliot *et al.*, 2021).

Questionnaire and behavioral studies of sex/gender differences in empathic ability

Early studies of sex/gender differences in empathic ability employing questionnaires (e.g. the Empathy Scale (Mehrabian and Epstein, 1972)) showed that women scored higher on self-reports of empathic ability in different sample sizes ranging from 20 to 600 (Eisenberg and Lennon, 1983). More recent research on sex/gender differences in empathic ability used other questionnaires such as the Empathy Quotient (Baron-Cohen and Wheelwright, 2004),

which requires self-reports of rating scores of 40 items such as 'I really enjoy caring for other people'. Studies using this questionnaire reported higher scores in women relative to men in either small (<100) or large (more than half a million) testing samples (Baron-Cohen and Wheelwright, 2004; Lawrence et al., 2004; Greenberg et al., 2018).

Researchers also examined sex/gender differences in different subdimensions of empathic ability using the Interpersonal Reactivity Index (IRI, Davis, 1983a), which consists of four subscales to assess distinct aspects of empathy. These include Perspective Taking (the ability to consider others' perspectives), Empathic Concern (the ability to feel warmth or compassion for others in deed), Fantasy (the ability to sit in a fictional situation) and Personal Distress (the ability to feel fear or anxiety in response to others' emotions). The IRI total score and the scores of the four subscales have been used to estimate people's general empathic ability and its distinct aspects. Similarly, previous studies using IRI reported higher scores in women than in men from different cultural samples (Davis, 1980; Eisenberg and Lennon, 1983; Thompson and Voyer, 2014). Sex/gender difference was observed either in all the four IRI subscales (e.g. Davis, 1980; Yang and Kang, 2020) or in some of the four IRI subscales (Gilet et al., 2013; Zhao et al., 2018) in different cultural samples. In addition, a study of adolescents aged between 13 and 16 years showed higher IRI scores in females than in males of the same age, and the sex/gender differences in IRI scores were increased with age (Mestre et al., 2009). Taken together, in general, questionnaire measures that depend on self-reports and reflect subjective estimations of empathic ability suggest women's superiority over men.

However, behavioral studies of the sex/gender difference in empathic ability have revealed incongruent results. When observing others' suffering shown in pictures and being asked to report the intensity of others' pain and own sad or upset feelings, women reported higher ratings compared to men but with a small effect size (Preis and Kroener-Herwig, 2012; Baez et al., 2017), possibly due to differences in general emotional responsiveness between women and men (Rueckert et al., 2011). Because empathy provides a psychological basis for prosocial behavior (Eisenberg and Fabes, 1990; Decety and Jackson, 2004; Hoffman, 2008), prosocial intentions have been measured in behavioral tests to infer sex/gender differences in empathic ability. The behavioral results did not always support women's superiority, and some studies even showed greater prosocial intentions in men, depending on social contexts such as recipients of help (Dorrough and Glöckner, 2019; Olsson et al., 2021). Prosocial behavior is sensitive to multiple factors including social goals/contexts besides empathic ability and thus may not provide an accurate estimation of empathic ability. A recent meta-analysis of 85 studies indicated that self-reported scores of empathic ability account for a negligible variance in behavioral cognitive empathy assessments, raising further concerns regarding the widespread use of self-reported measures as proxies for empathic ability and the relevant theoretical conclusions (Murphy and Lilienfeld, 2019).

Brain imaging studies of sex/gender differences in empathic ability

Since empathy is essentially a function of the brain, brain responses to others' emotional states (e.g. pain) can provide an objective estimation of empathic ability. The paradigm widely used to measure brain underpinnings of empathy is to record neural responses to perceived pain in others using various brain imaging techniques. Among these studies, functional magnetic

resonance imaging (fMRI) has been used to identify neural activities that are increased by perceived stimuli that induce suffering in others' body parts or painful facial expressions (e.g. Singer et al., 2004; Jackson and Decety, 2005; Lamm et al., 2007; Gu and Han, 2007; Sheng et al., 2014; Cui et al., 2015; Luo et al., 2015; see Lamm et al., 2011; Fan et al., 2011; Jauniaux et al., 2019; Ding et al., 2020; Fallon et al., 2020 for a review and meta-analysis). The neural networks identified as related to empathy for pain include the anterior/mid-cingulate and anterior insula (Singer et al., 2004; Jackson and Decety, 2005; Luo et al., 2015), sensorimotor cortex (Avenanti et al., 2005; Lamm et al., 2007; Riečanský and Lamm, 2019), supplementary motor area (Decety et al., 2008; Xu et al., 2009), temporoparietal junction (Lamm et al., 2007; Decety et al., 2008), medial prefrontal cortex (Mathur et al., 2010; Masten et al., 2011) and lateral frontal cortex (Gu and Han, 2007; Decety et al., 2008), indicating the validity of using fMRI as a measure of empathic ability. These brain regions have been associated with different components of empathy, such that anterior/mid-cingulate and anterior insula support affective sharing, the temporoparietal junction mediates self-other distinction when viewing others' suffering and the medial prefrontal cortex underlies inference of others' mental states and prosocial behavior (see Lamm et al., 2019 for a recent review). The mirror neuron system including the inferior frontal gyrus and inferior parietal lobule may also contribute to facilitating observers' abilities to understand and share others' emotional states (see Bekkali et al., 2021 for a systematic review and meta-analysis).

Despite the increasing number of fMRI studies of empathy, there has been no converging evidence for a reliable sex/gender difference in brain activities underlying empathy. An early meta-analysis of fMRI results failed to find greater neural responses to others' pain in the empathy network in women than in men (Lamm et al., 2011). Individual studies reported inconsistent results regarding sex/gender differences in empathic neural responses. For example, an fMRI study of 36 female and 34 male participants reported greater somatomotor responses to others' pain in women than in men (Christov-Moore and Iacoboni, 2019). Another fMRI study of 12 women and 12 men who were asked to infer the corresponding emotional expression of a masked face found stronger neural responses in the amygdala, hippocampus and superior temporal gyri in women but a greater activity in the temporoparietal junction in men (Derntl et al., 2010). Similarly, in a task focusing on one's own emotional response to emotion-expressing faces, an fMRI study of 14 female and 12 male participants showed stronger activations in the right inferior frontal cortex and superior temporal sulcus in women, whereas the temporoparietal junction was activated to a greater degree in men (Schulte-Rüther et al., 2008). Another fMRI study found greater sensitivity of empathic insular and cingulate responses to contextual modulations in men than in women (Singer et al., 2006). It appears that the conclusion regarding sex/gender differences in empathic brain activity is inconsistent and limited by the small samples tested in previous fMRI studies.

Electroencephalography (EEG) has been used to assess empathy by examining fast neural responses to perceived pain in others with a millisecond time resolution. There has been evidence that phase-locked EEG signals [i.e. event-related potentials (ERPs) that are both time locked and phase locked to stimulus onset] are modulated by perceived painful stimuli applied to others' body parts (Fan and Han, 2008; Han et al., 2008; Sessa et al., 2014; Cui et al., 2016, 2017) and painful facial expressions (Sheng and Han, 2012;

Sheng *et al.*, 2016; Palmieri *et al.*, 2020) as indexed by increased ERP amplitudes to painful compared to non-painful stimuli. Moreover, ERPs to others' pain within 200 ms after stimulus onset were much less influenced by top-down tasks compared to long-latency ERP components (Fan and Han, 2008; Sheng and Han, 2012). Non–phase-locked EEG signals (i.e. induced responses that are time locked but not phase locked to stimulus onset) also respond differentially to perceived painful and non-painful stimuli. Both EEG (Yang *et al.*, 2009; Perry *et al.*, 2010; Fabi and Leuthold, 2017; Joyal *et al.*, 2018; Li *et al.*, 2020; Riečanský *et al.*, 2020) and magnetoencephalography (MEG) (Cheng *et al.*, 2008; Whitmarsh *et al.*, 2011; Motoyama *et al.*, 2017) studies showed evidence that perceived noxious (*vs* innocuous) stimulations of others' body parts induced suppression (or desynchronization) of mu (7–12 Hz) and beta (13–30 Hz) band neural oscillations. There were also MEG and EEG findings that perceived painful (*vs* non-painful) stimulations to others' body parts enhanced alpha oscillations (Mu *et al.*, 2008; Levy *et al.*, 2016). MEG research localized desynchronization of mu rhythm in response to painful (*vs* non-painful) stimulations to others' body parts in the somatosensory cortex (Cheng *et al.*, 2008; Whitmarsh *et al.*, 2011; Motoyama *et al.*, 2017), supporting a key role of sensorimotor processes in empathy for others' pain (Riečanský and Lamm, 2019). Perceived painful (*vs* non-painful) expressions of faces, however, were associated with increased alpha oscillations in the precuneus/parietal cortices followed by increased alpha-band oscillations in the left anterior insula and temporoparietal junction (Zhou and Han, 2021). Most importantly, EEG research has shown evidence that both phase-locked EEG and non–phase-locked EEG signals in response to perceived painful stimulation applied to others or others' painful expressions predict subjective evaluations of others' pain or one's own unpleasantness induced by others' pain (Fan and Han, 2008; Mu *et al.*, 2008; Sheng and Han, 2012). These results suggest the validity of the EEG/MEG methods for assessing individuals' empathic abilities.

Similar to the results of fMRI studies, EEG/MEG research has not shown reliable evidence for sex/gender effects on neural responses to perceived pain in others. An EEG study of 26 healthy adults (13 women and 13 men) found that both males and females showed greater neural responses to pictures of hands in painful relative to non-painful conditions at 140 ms after stimulus onset over the frontal lobe and after 380 ms over the central-parietal regions (Han *et al.*, 2008). However, the amplitudes of differential neural responses to perceived painful (*vs* non-painful) stimulations to others did not differ significantly between female and male participants although long-latency empathic responses to others' pain were modulated by task demands, which required attention to or distracted attention away from the pain cues, to a larger degree in women than in men. An MEG study of 32 adults (16 women and 16 men) observed stronger suppressions of 10 Hz neural responses to both painful and non-painful stimuli in women than in men but did not report any sex/gender difference in differential neural responses to painful *vs* non-painful stimuli (Yang *et al.*, 2009). Together, the results of brain imaging studies of empathy were usually obtained from small samples and provided no evidence for women's superiority in empathy-specific brain activities.

The contradictory results reported in the previous questionnaire/behavioral and brain imaging studies of sex/gender differences in empathic ability appear to be hard to explain. It is difficult to compare the results of the questionnaire and brain imaging measures that were reported in separate studies with different sample sizes. The small sample size in brain imaging

studies due to high experimental costs and long experimental time did not allow the exclusion of the effect of individual differences in empathic brain activity when comparing the mean empathic neural responses between women and men. To clarify the inconsistent results of the previous questionnaire and brain imaging studies requires comparisons of questionnaire and brain imaging measures of empathic ability by testing the same cohort with a reasonable sample size. Moreover, previous brain imaging studies usually reported the absence of a significant sex/gender group difference in empathic brain activities based on a threshold defined as a specific *P*-value. Logically and statistically, a failure to find a significant sex/gender difference in empathic brain activities does not necessarily reach a reliable conclusion of no sex/gender difference in empathic ability. Such a conclusion needs to be tested by reasonable statistical analyses. Importantly, a theoretical account is required to integrate the seemingly contradictory results regarding sex/gender differences in empathy reported by the previous questionnaire and brain imaging studies.

The present research

The current work sought to investigate sex/gender differences in empathic ability by overcoming both methodological and theoretical challenges faced by previous research. In Study 1, we examined sex/gender differences in empathic ability based on self-reported questionnaire measures in a relatively large sample. We chose IRI to estimate empathic ability so as to assess sex/gender differences in different aspects of empathy separately. In Study 2, we collected both IRI and EEG measures of empathic ability. EEG measures were selected because EEG signals originate directly from neural responses, have a high time resolution and are less easy to be consciously controlled relative to rating scores of questionnaire measures. In Study 2, we collected both IRI rating scores and EEG signals from a sample (141 women and 145 men) that was larger than previous brain imaging studies of empathy. We analyzed both phase-locked and non–phase-locked EEG signals, which are sensitive to painful (*vs* neutral) expressions and correlated with one's own emotions induced by viewing others' pain (Sheng and Han, 2012; Sheng *et al.*, 2016), as objective estimations of sex/gender differences in empathic ability. Unlike previous research, we analyzed EEG results by conducting Bayesian analyses to test a null hypothesis regarding the sex/gender difference in brain responses to others' pain.

Given that the results of Studies 1 and 2 confirmed a reliable sex/gender difference in empathic ability in questionnaire measures but a null effect in EEG measures, in Study 3, we further tested a hypothesis that social expectations contribute to the observed sex/gender difference in questionnaire measures of empathic ability. Early research suggested a motivational interpretation of sex/gender differences in empathy-related measures. This interpretation was supported by an early meta-analysis work (Ickes *et al.*, 2000) and empirical findings that sex/gender differences in empathy-related measures were observed only when empathy-relevant gender-role expectations, obligations or awareness of being evaluated were made salient (Berman, 1980; Eisenberg and Lennon, 1983; Klein and Hodges, 2001). A recent work also showed that sex/gender differences in self-reports of empathic capacity, estimated using the Empathy Quotient (Baron-Cohen and Wheelwright, 2004), increased when participants were explicitly informed that their empathic capacity would be assessed (Löffler and Greitemeyer, 2021). In Study 3, we tested a causal relationship between social expectations of women or

men's empathic ability and sex/gender differences in empathic ability estimated using the IRI. We collected IRI rating scores from participants after they had been primed with information that reminds social expectations that women (or men) are good at sharing and caring for the feelings of others (an empathy-inducing condition) or information that women (or men) are independent and good at self-regulation (a control condition). We tested whether the sex/gender difference in IRI rating scores appears in the control condition but not in the empathy-inducing condition.

Together the results of the three studies in the current work showed evidence for the dissociation between subjective (questionnaire) and objective (EEG) measures in the estimation of sex/gender differences in empathic ability. In addition, our results suggest a causal role of social expectations in generating sex/gender differences in questionnaire estimations of empathic ability. Our findings indicate that whether the notion of women's superiority in empathy reflects a biological difference between the two sexes or a gender-role stereotype remains an open question.

Study 1: Questionnaire estimation of sex/gender differences in empathic ability

In Study 1, we sought to conduct questionnaire estimations of sex/gender differences in empathic ability by testing a large Chinese sample with IRI. We compared IRI total score to assess sex/gender differences in self-report of empathic ability as a general construct. Although some studies on empathy combined the four IRI subscales to form two factors (e.g. combine the Perspective Taking and the Fantasy subscales into a 'Cognitive Empathy' factor, and the Empathic Concern and the Personal Distress subscales into an 'Affective Empathy' factor), research using the confirmatory factor analysis showed evidence for better model fit for the four-factor than two-factor structure (see Chrysikou and Thompson (2016) for literature reviews and results of the confirmatory factor analysis). Therefore, we further compared the scores of each IRI subscale between male and female participants. We expected higher rating scores in women than in men, given the results of previous studies of empathic ability across different cultural samples (Davis, 1980; Eisenberg and Lennon, 1983; Thompson and Voyer, 2014). In Study 1, we reanalyzed the data from a Chinese sample in our previous research (Luo *et al.*, 2015) and reported results of sex/gender differences in IRI scores that have not been reported before.

Methods

In Study 1, we recruited 1486 Chinese college students from a Chinese sample in our previous research (Luo *et al.*, 2015), whose gender/age information and IRI scores were completed (790 men, mean age $\pm$ s.d. $= 18.8 \pm 1.9$ years; 696 women, mean age $\pm$ s.d. $= 18.8 \pm 1.8$ years). The Chinese version of the IRI questionnaire that was validated in previous research (Huang *et al.*, 2012) was used in the current study. The questionnaire consists of four 5-point subscales (0 = does not describe me well, 4 = describe me very well) that assess different aspects of empathic ability, including Perspective Taking, Fantasy, Empathic Concern, and Personal Distress, and was printed on a piece of paper. The experimental protocols in all studies were approved by the local ethics committee at the School of Psychological and Cognitive Sciences of Peking University. All participants were paid for their participation.

Results and discussion

To assess sex/gender differences in subjective estimations of empathic ability, we analyzed the rating scores of the IRI questionnaire from 696 women and 790 men in Study 1. Rating scores were subject to independent two-sample t-tests, and the results revealed significantly higher total IRI scores in women than in men ($t(1484) = 5.874$, $P < 0.001$, False Discovery Rate [FDR] corrected, Cohen's $d = 0.305$, 95% confidence interval [CI] $= [0.203–0.408]$, Figure 1). Women, compared to men, also scored significantly higher on the subscales of Personal Distress ($t(1484) = 8.102$, $P < 0.001$, FDR corrected, Cohen's $d = 0.421$, 95% CI $= [0.318–0.524]$), Empathic Concern ($t(1484) = 2.299$, $P = 0.028$, FDR corrected, Cohen's $d = 0.120$, 95% CI $= [0.018–0.221]$) and Fantasy ($t(1484) = 5.090$, $P < 0.001$, FDR corrected, Cohen's $d = 0.265$, 95% CI $= [0.162–0.367]$) but not on the subscale of Perspective Taking ($t(1484) = -0.870$, $P = 0.384$, FDR corrected, Cohen's $d = -0.045$, 95% CI $= [-0.147–0.057]$). These results indicate that, relative to men, women reported higher IRI total scores and scores of the Fantasy, Empathic Concern and Personal Distress subscales. Although the effect sizes of sex/gender differences in the rating scores were small, the pattern of women's superiority over men was consistently observed for IRI total score and the IRI subscales (except the Perspective Taking subscale). In addition, the results of our work that tested a Chinese sample are consistent with the results of previous IRI-based studies that tested samples in North America and Europe (Davis, 1980; Eisenberg and Lennon, 1983; Thompson and Voyer, 2014). The findings of previous and current studies together indicate that IRI estimations of empathic ability suggest women's superiority over men regardless of different cultural samples tested in independent studies.

Study 2: Questionnaire and EEG measures of sex/gender differences in empathic ability

In Study 2, we examined whether both subjective and objective estimations of sex/gender differences in empathic ability support women's superiority over men by collecting both questionnaire and EEG measures of empathic ability from an independent Chinese sample. We collected the IRI questionnaire prior to EEG recording and expected replication of the results in Study 1. We recorded EEG signals from participants while they made pain *vs* no-pain judgments on rapid presentation of faces with painful or neutral expressions. We analyzed both phase-locked and non–phase-locked neural responses to painful (*vs* neutral) expressions as objective estimations of empathic ability similar to previous research (Sheng and Han, 2012; Sheng *et al.*, 2016). We also conducted Bayesian analyses of empathic neural responses to test the null hypothesis regarding the sex/gender difference in brain responses to others' pain.

Methods
Participants

The sample size in Study 2 was pre-determined based on G*Power estimation (Faul *et al.*, 2009). A sample size of 266 participants was required to obtain a small effect size of 0.1, with an error probability of 0.05 and power of 0.90 for a within-between interaction of a repeated-measures analysis of variance (ANOVA). The results suggested a sample size of 266 for observing a reliable sex/gender difference in questionnaire measures of empathic ability. We recruited 286 Chinese college students (145 men, mean age $\pm$ s.d. $= 20.8 \pm 1.0$ years; 141 women, mean age $\pm$ s.d. $= 20.4 \pm 0.8$ years). This sample size, being much larger
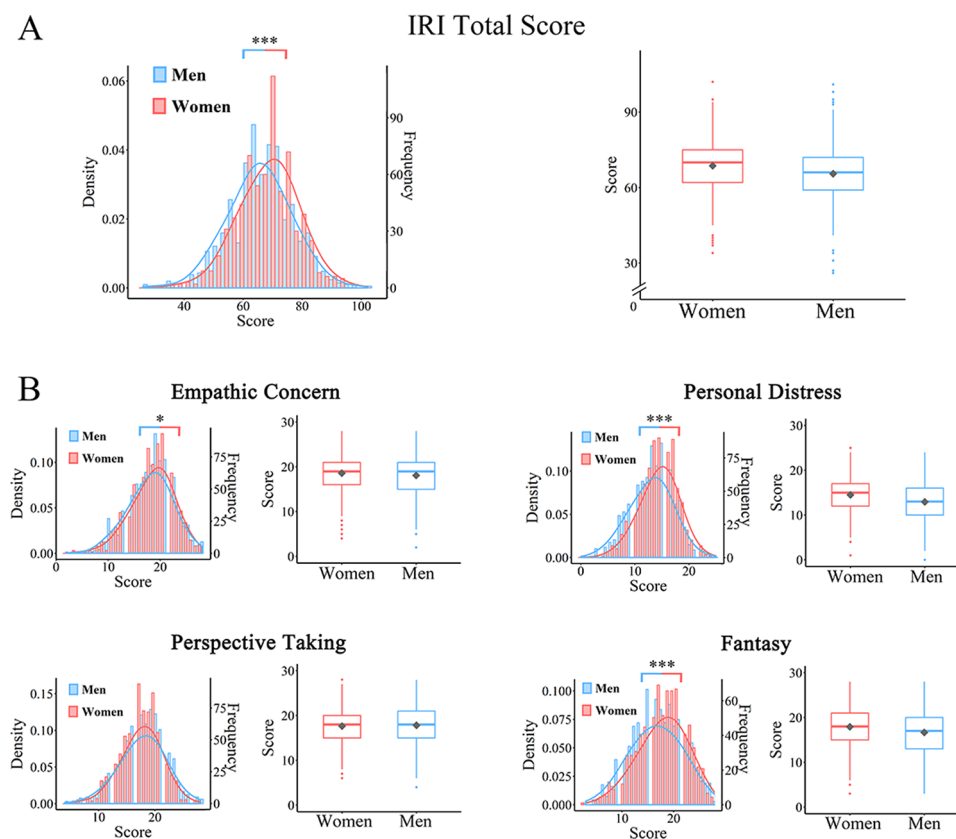
Fig. 1. Results of questionnaire estimation of sex/gender differences in empathic ability in Study 1. (A) The density distribution and mean of the total IRI score. (B) The illustration of the density distribution and mean scores of each IRI subscale. *P < 0.05; ***P < 0.001. The left panel shows the density (the left y-axis) and frequency (the right y-axis). Kernel density estimation was conducted to assess the probability density function of the questionnaire measures. The right panel illustrates bar charts of the total IRI score. The lower and upper hinges of the boxes correspond to the 25th and 75th percentiles, respectively. Each dot represents an outlier. The horizontal line inside each box shows the median. The lower and higher whiskers represent the lowest and highest observed values excluding outliers, respectively. The gray diamond in the middle represents the mean of the data.

than those in our previous EEG research (e.g. Sheng and Han, 2012), was large enough to reveal robust neural responses to painful *vs* neutral expressions with a great statistical power (e.g. observed power for phase-locked interaction result is 0.95) and a small estimation error (Ioannidis, 2005). The analysis of questionnaire results included all participants. Fifty participants were excluded from EEG data analyses due to lack of behavioral data (15 participants), EEG recording errors (3 participants), no response during EEG recording (3 participants), EEG artifact rejection (<50% trials after artifact rejection, 12 participants) or being identified as outliers (defined by 3 s.d. away from the mean of the normalized empathic neural responses, 17 participants). There were 236 participants left for further statistical analyses (118 men, mean age ± s.d. = 20.8 ± 1.0 years; 118 women, mean age ± s.d. = 20.5 ± 0.8 years). All participants were right-handed (except one male participant was left-handed) and had normal or corrected-to-normal vision. All participants reported no history of neurological or psychiatric diagnoses or medication, drug, or alcohol abuse. To examine whether the menstrual cycle phase is associated with questionnaire and EEG measures of empathy, we asked female participants to report their menstrual cycle phase (days of the menstrual cycle). Participants provided written informed consent after the experimental procedure had been fully explained and were informed of their right to withdraw at any time during the study. All participants were paid for their participation.

Stimuli and procedure

Before EEG recording, all participants completed the IRI questionnaire (Davis, 1983b) printed on a piece of paper. The stimuli used during EEG recording were adopted from the previous work (Sheng and Han, 2012), which consisted of 16 Asian faces (8 men) with neutral or painful expressions. Each face was displayed in the center on a gray background and subtended a visual angle of 3.8° × 4.7° at a viewing distance of 60 cm. On each trial, a face was presented for 200 ms, followed by a fixation cross with a duration varying randomly between 800 and 1400 ms. The participants were asked to identify painful *vs* neutral expressions by pressing one of the two keys on a keyboard using the left and right index fingers. There were 8 blocks of 32 trials. In each block, each of the 16 faces was presented twice with painful or neutral expressions in a random order.

EEG data acquisition and analysis

EEG was continuously recorded from 32 scalp electrodes using the NeuroScan system (Curry 7, Compumedics Neuroscan, TX) and was re-referenced to the average of the left and right mastoid electrodes offline. Electrode impedance was kept <5 kΩ. Two electrodes located above and below the left eye were used to monitor eye blinks and vertical eye movements. The horizontal electrooculogram was recorded from electrodes placed 1.5 cm lateral to the left and right external canthi. EEG signals

were amplified with an online band-pass filter of 0.01–400 Hz and digitized at a sampling rate of 1000 Hz. EEG data were filtered with a low-pass filter at 40 Hz and a high-pass filter at 0.5 Hz offline. Artifacts related to eye movement or eye blinks were removed using the independent component analysis analysis implemented in the EEGLAB toolbox. ERPs in each condition were averaged separately offline with an epoch beginning 200 ms before stimulus onset and continuing for 1000 ms. Trials contaminated by eye blinks, eye movements or muscle potentials exceeding $\pm 100 \mu V$ at any electrodes were excluded from the average. This criterion was determined to leave enough number of trials for average analyses. There were 122 ± 9 trials accepted for average in each condition after artifact rejection. The baseline for ERP measurements was the mean voltage of a 200 ms pre-stimulus interval, and the latency was measured relative to stimulus onset.

To test sex/gender differences in phase-locked empathic responses, we quantified phase-locked neural responses in each condition by calculating the mean amplitude of each ERP component including the N1, P2, N2 and long-latency positivity (LPP) at the frontal/central electrodes (Fz, FCz, Cz, FC3 and FC4) and the mean values of amplitudes of the P1 and N170 at the right occipitotemporal electrode (T6). The time window for calculations of the mean amplitude of each ERP component was determined based on ERP waveforms and topographic maps collapsed across conditions. The mean amplitudes were then subject to ANOVAs with Expression (painful *vs* neutral) as a within-subjects variable and Gender (women *vs* men) as a between-subjects variable. According to Steiger (2004), the general rule of thumb to use the CIs to test a statistical hypothesis (H0) is as follows: when testing a two-sided/one-sided hypothesis at the alpha level, use a $100 \times (1 - \alpha)\%/100 \times (1–2\alpha)\%$ CI. We reported 90% CIs of η2 in ANOVA test because the hypothesis test is one-sided that η2 cannot be negative. This hypothesis test is equivalent to the standard ANOVA *F* test. Given the previous findings of sex/gender differences in brain structures (Luders and Toga, 2010; Ruigrok *et al.*, 2014; Liu *et al.*, 2020) and in empathy-irrelevant ERP components to faces (e.g. Sun *et al.*, 2010), we defined normalized empathic neural responses as Mean Amplitude$_{\text{(painful—neutral expressions)}}$/Mean Amplitude$_{\text{(painful + neutral expressions)}}$ of each ERP component. A comparison of the normalized empathic neural responses between women and men allowed us to examine possible sex/gender differences of empathic neural responses by controlling sex/gender difference in brain activity that is unrelated to empathy.

We also examined sex/gender differences in non–phase-locked empathic responses at the frontal/central electrodes (Fz, FCz, Cz, FC3 and FC4). We first quantified non–phase-locked neural responses to painful and neutral faces by conducting band-pass filtering of EEG data (a low pass: 80 Hz; a high pass: 0.5 Hz). EEG results in each condition were averaged separately offline, with an epoch beginning 400 ms before stimulus onset and continuing for 1200 ms. ERPs in each condition were subtracted from corresponding EEG epochs to remove phase-locked EEG activities. The spectra power of trials in the same condition was averaged to obtain non–phase-locked responses.

Spectra powers of neural oscillations were quantified based on a wavelet decomposition of the signal between 1 and 40 Hz in 1 Hz steps, given that most EEG and MEG research observed non–phased-locked empathic responses lower than 40 Hz (e.g. Cheng *et al.*, 2008; Mu, *et al.*, 2008 ; Zhou and Han, 2021). The signal was then convoluted by the complex Morlet wavelet $w(t, f_0)$ (Kronland-Martinet *et al.*, 1987), with a Gaussian shape in time (s.d. σ_t) and

frequency (s.d. σ_f) domains around its central frequency f_0:

$$\mathbf{w}(\mathbf{t}, f_0) = A e^{-t^2/(2\sigma_t^2)} e^{2i\pi f_0 t}$$

with $\sigma_f = 1/2\pi\sigma_t$. Wavelets were normalized so that their total energy is 1, and the normalization factor A being equal to $(\sigma t \sqrt{\pi})^{-1/2}$. A wavelet family is characterized by a constant ratio (f_0/σ_f), which should be chosen in practice greater than ~ 5 (Grossmann *et al.*, 1990). The wavelet family was defined by $f_0/\sigma_f = 5$ (wavelet duration $2\sigma t$ of ~ 1.6 periods of oscillatory activity at f_0), with f_0 ranging from 1 to 40 Hz in 1 Hz steps. The mean time-frerquency (TF) energy in a pre-stimulus window (–300 to –100 ms) calculated as the baseline power was subtracted from the post-stimulus (0 to 1200 ms) TF power in each frequency band and was then subjected for further statistical analysis.

Similar to the previous research (Pfurtscheller and Lopes da Silva, 1999), we calculated the percentage increase/decrease of time–frequency power in the following way: Event-related synchronization/Event-related desynchronization $=$ [(A – R)/R] $\times 100\%$, where A refers to the spectrum power in a specific time window post-stimulus and R refers to the spectrum power in the pre-stimulus 200 ms window (–300 to –100 ms). Time–frequency power values at each time point and each frequency band in different situations (painful *vs* neutral stimuli) were then subjected to paired *t*-tests to test the main effect of painful expression. Similarly, we calculated normalized non–phase-locked empathic responses as Mean Power$_{\text{(painful—neutral expressions)}}$/Mean Power$_{\text{(painful + neutral expressions)}}$, which were further compared between female and male participants. In addition, normalized non–phase-locked empathic responses at each time point and each frequency were subject to independent *t*-test to examine interactions. All results were subject to FDR correction for multiple comparisons.

Because the normalized EEG data set did not meet the assumptions of classical parametric tests, we conducted non-parametric bootstrap analyses (Efron and Tibshirani, 1994) to assess the probability of observing differences in empathic neural responses between female and male participants. The bootstrapping procedure included (i) creating a bootstrapping female data set by randomly selecting 118 participants with replacement from the female group; (ii) creating a bootstrapping male data set by randomly selecting 118 participants with replacement from the male group; and (iii) calculating a *t*-value of the difference between the two bootstrapping data sets. After 10 000 iterations of this procedure, we obtained the distribution of the *t*-values. The observed *t*-value obtained by comparing the original male and female samples was then calculated and compared along the permutated distribution of *t*-values. The difference between female and male groups was considered to be significant only if the probability of the observed *t*-value along the permutated distribution of *t*-value is <5% (two tailed).

To further verify the null results obtained in the bootstrap analyses, we conducted the Bayes factor (BF) analysis (Dienes, 2011) to assess the likelihood of data-driven alternative hypothesis over the likelihood of data-driven null hypothesis (BF10) or the reverse (BF01) with default priors for paired *t*-test design (a default prior rscale of sqrt(2)/2 = 0.707 was used, corresponding to a medium effect size). Compared with widely used *P*-values, BFs allow researchers to make a statement to support a null hypothesis and to estimate the amount of evidence present in the data. Using R (version 3.5.1) with a 'BayesFactor' Package (Morey and Rouder, 2018), we calculated BF10, which represents the ratio that contrasts the likelihood of the data fitting under the alternative
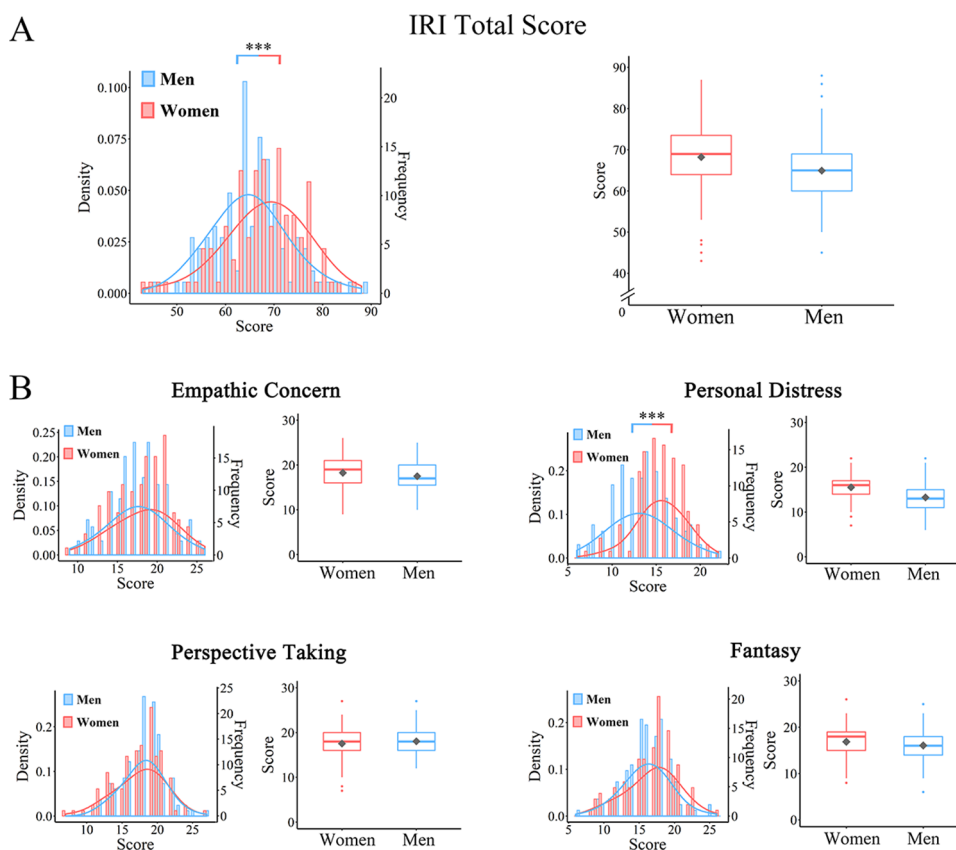
Fig. 2. Results of questionnaire estimation of sex/gender differences in empathic ability in Study 2. (A) The density distribution and mean of the total IRI score. (B) The illustration of the density distribution and mean scores of each IRI subscale. ***$P < 0.001$. The left panel shows the density (the left y-axis) and frequency (the right y-axis). Kernel density estimation was conducted to assess the probability density function of the questionnaire measures. The right panel illustrates bar charts of the total IRI score. The lower and upper hinges of the boxes correspond to the 25th and 75th percentiles, respectively. Each dot represents an outlier. The horizontal line inside each box shows the median. The lower and higher whiskers represent the lowest and highest observed values excluding outliers, respectively. The gray diamond in the middle represents the mean of the data.

hypothesis with the likelihood of the data fitting under the null hypothesis. A BF10 <1 is regarded as supporting the null hypothesis, whereas a BF10 >1 is regarded as supporting the alternative hypothesis.

Results and discussion
Results of questionnaire measures

Independent two-sample *t*-tests of the questionnaire rating scores showed that women (*vs* men) reported significantly higher IRI total scores ($t(234) = 3.541$, $P < 0.001$, FDR corrected, Cohen's $d = 0.460$, 95% CI = [0.202–0.719], Figure 2). Women (*vs* men) also scored significantly higher on the subscales of Personal Distress ($t(234) = 5.568$, $P < 0.001$, FDR corrected, Cohen's $d = 0.725$, 95% CI = [0.461–0.988]) but not on the subscale of Perspective Taking ($t(234) = -1.115$, $P = 0.266$, FDR corrected, Cohen's $d = -0.145$, 95% CI = [−0.401 to 0.110]), Fantasy ($t(234) = 2.026$, $P = 0.065$, FDR corrected, Cohen's $d = 0.264$, 95% CI = [0.007–0.520]) and Empathic Concern ($t(234) = 1.954$, $P = 0.065$, FDR corrected, Cohen's $d = 0.254$, 95% CI = [0.002–0.509]). These results replicated the results in Study 1 and showed further evidence that IRI estimations of empathic ability suggest women's superiority over men.

Behavioral results of the pain judgment task during EEG recording

Table 1 shows response accuracies and reaction times during pain judgments on painful and neutral expressions of faces during EEG

Table 1. Reaction times and response accuracies (mean ± s.d.) during pain judgments in Study 2

	Female participants		Male participants	
	Painful faces	**Neutral faces**	**Painful faces**	**Neutral faces**
Reation Time (ms)	530 ± 58	560 ± 66	543 ± 65	561 ± 61
Accuracy (%)	94 ± 3.8	93 ± 5.3	93 ± 4.7	93 ± 4.8

recording. To avoid the effect of speed-accuracy trade-off on comparisons of behavioral performances of women and men, we calculated performance efficiency (defined as reaction times divided by response accuracy). The ANOVA of performance efficiencies with Expression (painful *vs* neutral) as a within-subjects variable and Gender (women *vs* men) as a between-subjects variable showed a significant main effect of Expression ($F(1234) = 65.968$, $P < 0.001$, $\eta_p^2 = 0.220$, 90% CI = [0.146–0.292]) and a significant two-way interaction ($F(1234) = 8.910$, $P = 0.003$, $\eta_p^2 = 0.037$, 90% CI = [0.007–0.084]). Simple effect analyses revealed that women performed better than men in the painful condition ($F(1234) = 6.789$, $P = 0.010$, $\eta_p^2 = 0.028$, 90% CI = [0.004–0.071]) but not in the neutral condition ($F(1234) = 0.083$, $P = 0.774$, $\eta_p^2 < 0.001$, 90% CI = [0.000–0.013]), suggesting faster reactions with higher accuracies in response to painful expressions in women.
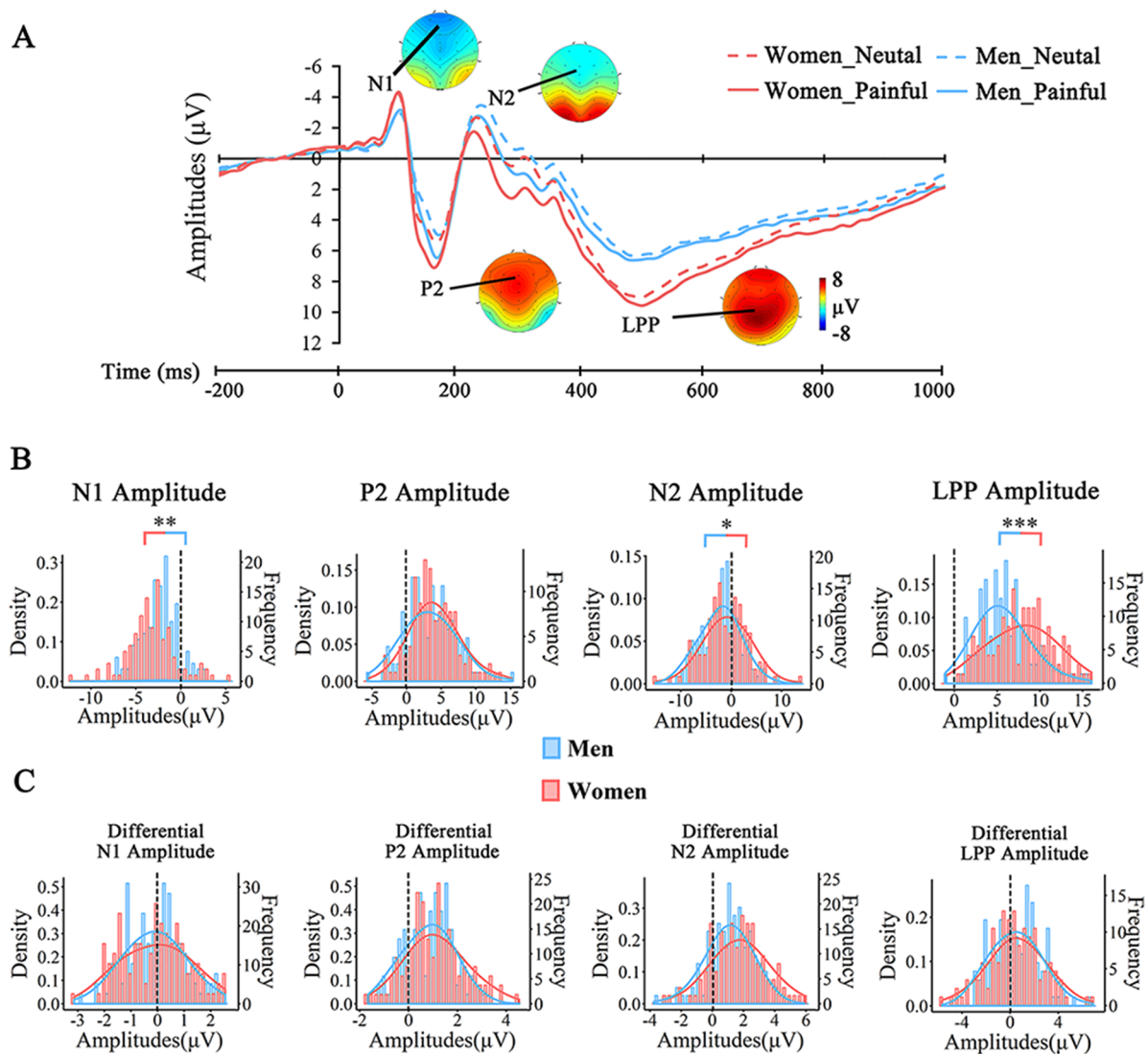
Fig. 3. ERP results at the frontal/central electrodes in Study 2. (A) The illustration of ERPs to painful and neutral faces in women and men. (B) The density (the left y-axis) and frequency (the right y-axis) distributions of the N1/P2/N2/LPP amplitudes in responses to face stimuli in women and men. (C) The density (the left y-axis) and frequency (the right y-axis) distributions of the differential N1/P2/N2/LPP amplitudes in response to painful (*vs* neutral) faces in women and men. Kernel density estimation was conducted to assess the probability density function of the mean amplitudes of each ERP component and difference wave. *P < 0.05; **P < 0.01; ***P < 0.001.

## Results of phase-locked empathic neural responses

We first analyzed phase-locked neural responses to perceived painful (*vs* neutral) expressions of faces as an objective index of empathic ability, which were then used to estimate sex/gender differences in empathic ability. ERPs to faces over the frontal-central regions were characterized by an early negative wave at 70–110 ms (N1) and a positive wave at 140–180 ms (P2) followed by a negative wave at 220–270 ms (N2) and an LPP at 400–550 ms (Figure 3). Face stimuli also elicited a positivity at 60–130 ms (P1) and a negativity at 140–180 ms (N170) at the lateral occipitotemporal electrodes (Figure 4). The mean amplitude was extracted around the peak of each ERP component by which we calculated normalized empathic neural responses as Mean Amplitude$_{(painful—neutral expressions)}$/Mean Amplitude$_{(painful + neutral expressions)}$ of each ERP component. The normalized empathic neural responses ruled out influences of sex/gender differences in neural activities that were irrelevant

to empathy on cross-sex/gender comparison of objective estimations of empathic ability.

Results of ANOVAs of ERP amplitudes. ANOVAs of the N1 amplitudes showed only a significant main effect of Gender (F(1234) = 10.800, P = 0.001, $\eta_p^2 = 0.044$, 90% CI = [0.011–0.094]) due to a larger (more negative) N1 amplitude in women than in men (Figure 3). ANOVAs of the P2 amplitudes revealed a significant main effect of Expression (F(1234) = 264.797, P < 0.001, $\eta_p^2 = 0.530$, 90% CI = [0.461–0.587]) as the P2 amplitude was enlarged by painful than by neutral expressions (Figure 3).

ANOVAs of the N2 amplitudes showed significant main effects of Expression (F(1234) = 138.467, P < 0.001, $\eta_p^2 = 0.372$, 90% CI = [0.293–0.440]) and Gender (F(1234) = 5.023, P = 0.026, $\eta_p^2 = 0.021$, 90% CI = [0.001–0.060]), suggesting that the N2 amplitudes were sensitive to painful *vs* neutral expressions and
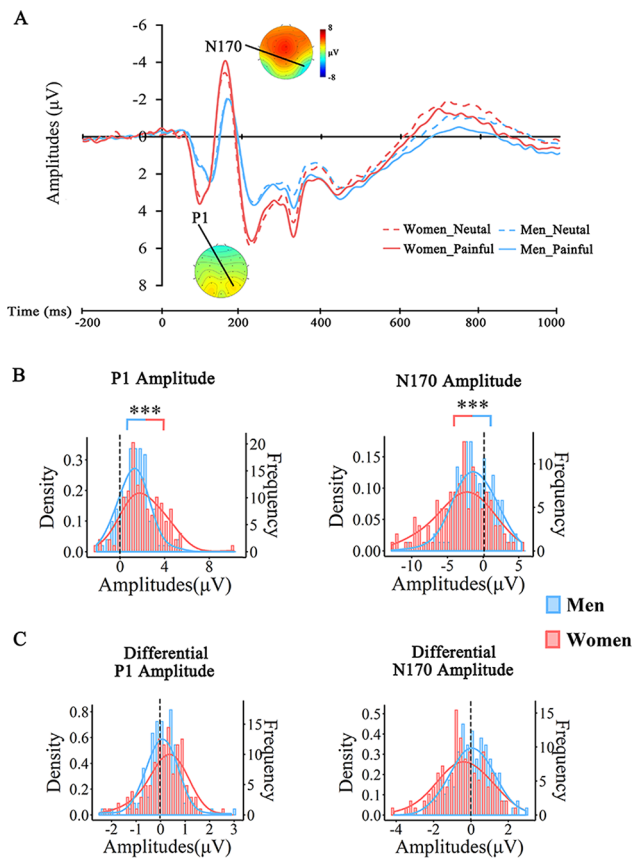
Fig. 4. ERP results at the right occipitotemporal electrode in Study 2. (A) The illustration of ERPs to painful and neutral faces at electrode T6 where N170 showed the largest amplitude among all other electrodes. (B) The density (the left y-axis) and frequency (the right y-axis) distributions of the P1/N170 amplitudes in responses to face stimuli in women and men. (C) The density (the left y-axis) and frequency (the right y-axis) distributions of the differential P1/N170 amplitudes in response to painful (*vs* neutral) faces in women and men. Kernel density estimation was conducted to assess the probability density function of the mean amplitudes of each ERP component and difference wave. ***$P < 0.001$.

were larger in men (*vs* women, Figure 3). ANOVAs of the LPP amplitudes showed significant main effects of Expression ($F(1234) = 15.460$, $P < 0.001$, $\eta_p^2 = 0.062$, 90% CI = [0.021–0.117]) and Gender ($F(1234) = 32.040$, $P < 0.001$, $\eta_p^2 = 0.120$, 90% CI = [0.062–0.186]), indicating larger LPP amplitudes to painful than to neutral expression and in women than in men.

ANOVAs of the P1 amplitudes showed a significant main effect of Gender ($F(1234) = 19.716$, $P < 0.001$, $\eta_p^2 = 0.078$, 90% CI = [0.031–0.137]) due to larger P1 amplitudes in women (*vs* men) (Figure 4). There was also a significant main effect of Expression ($F(1234) = 6.980$, $P = 0.008$, $\eta_p^2 = 0.029$, 90% CI = [0.004– 0.073]) though the effect size was small. ANOVAs of the N170 amplitudes showed only a significant main effect of Gender ($F(1234) = 20.706$, $P < 0.001$, $\eta_p^2 = 0.081$, 90% CI = [0.034– 0.141]) due to larger N170 amplitude in women than in men (Figure 4).

Results of bootstrap analyses of normalized ERP amplitudes.
We calculated normalized P2 empathic responses and conducted a non-parametric bootstrap analysis to examine potential sex/gender difference. However, the results failed to find a significant effect of Gender ($t(234) = -1.981$, $P = 0.053$, Cohen's $d = 0.258$, 95% CI = [0.001–0.514]). We further conducted a Bayes factor analysis

of the normalized P2 empathic responses to compare the alternative hypothesis of greater empathic neural responses in women (*vs* men) and the null hypothesis of no sex/gender difference. This analysis revealed a BF10 of 0.900 that supports the null hypothesis.

A non-parametric bootstrap analysis did not show a significant difference in normalized N2 empathic responses to painful (*vs* neutral) expressions between women and men ($t(234) = -0.217$, $P = 0.835$, Cohen's $d = -0.028$, 95% CI = [−0.283–0.227]). This is further reinforced by the result of the Bayes factor analysis that revealed a Bayes factor of BF10 = 0.146 that supports the null hypothesis. Similarly, a non-parametric bootstrap analysis failed to show evidence for significant sex/gender difference in normalized LPP responses to painful (*vs* neutral) expressions ($t(234) = -0.408$, $P = 0.686$, Cohen's $d = -0.053$, 95% CI = [−0.308–0.202]). The results of the Bayes factor analysis (BF10 = 0.154) also support the null hypothesis regarding sex/gender difference.

A non-parametric bootstrap analysis did not show a significant sex/gender difference in normalized P1 neural responses to painful (*vs* neutral) expressions ($t(234) = 0.074$, $P = 0.953$, Cohen's $d = 0.010$, 95% CI = [−0.246–0.256]). The results of the Bayes factor analysis (BF10 = 0.143) also support the null hypothesis regarding sex/gender difference.

Results of non–phase-locked empathic neural responses
We further estimated sex/gender differences in non–phase-locked empathic neural responses to painful (*vs* neutral) faces by conducting time–frequency analyses of EEG signals. Time–frequency power values in each time point and each frequency in painful and neutral face conditions were first subject to paired *t*-tests to identify neural oscillations with greater power in women than in men and in response to painful than neutral faces. The result identified a cluster of interest (4–6 Hz, 133–499 ms, at FT7, T3, TP7 and T5) in which the time–frequency power was significantly larger in response to painful compared to neutral faces ($P < 0.05$, FDR correction, Figure 5A). The results also revealed a significant cluster of interest (15–27 Hz, 92–497 ms, at T6 and TP8) in which the time–frequency power was larger in women than in men ($P < 0.05$, FDR correction, Figure 5B). Time–frequency power values were then extracted from this cluster showing the main effect of facial expressions to calculated normalized empathic responses for examination of potential sex/gender differences. However, a non-parametric bootstrap analysis of the normalized empathic responses failed to show a significant effect ($t(234) = 1.023$, $P = 0.445$, Cohen's $d = 0.133$, 95% CI = [−0.122–0.388], Figure 5C), providing no evidence for a reliable sex/gender difference. Similarly, the results of a Bayes factor analysis of the normalized empathic responses (BF10 = 0.233) support the null hypothesis regarding sex/gender difference. We also performed independent *t*-tests of differential time–frequency power values (painful minus neutral faces) in each time point and each frequency between women and men but failed to find any significant sex/gender difference ($P < 0.05$, FDR correction, Figure 5D).

Correlation between IRI scores and empathic neural responses
Given that questionnaire and EEG measures showed distinct results regarding sex/gender differences in empathic ability, we further examined whether questionnaire and EEG measures of empathic ability covaried across all participants as one sample.
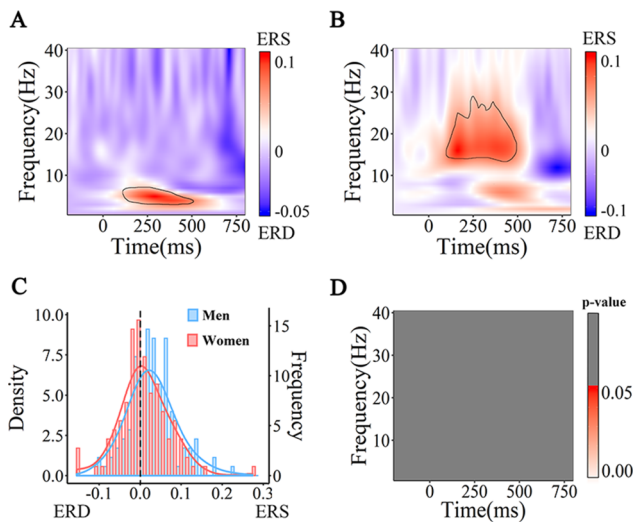
Fig. 5. The results of time–frequency power analyses of empathic neural responses. (A) The illustration of time–frequency power to painful *vs* neutral faces at electrode T5. The time–frequency power of neural responses to faces in the cluster encircled was larger in response to painful compared to neutral faces. (B) The illustration of the sex/gender difference in time–frequency power unrelated to empathic ability at electrode T6. The time–frequency power in the cluster encircled was larger in women than in men. (C) The density (the left y-axis) and frequency (the right y-axis) plots used a non-parametric way (kernel density estimation) to estimate the probability density function of the mean cluster of interest of empathy main effect extracted as above. (D) The illustration of statistical significance regarding sex/gender difference in non–phase-locked empathic neural responses.

We conducted correlation analyses of the relationships between IRI scores (total score and subscale scores) and phase-locked/non–phase-locked empathic neural responses. The results did not show any significant correlation ($Ps < 0.05$, FDR corrected, Table S1 for details), suggesting a disassociation of questionnaire and EEG measures of empathic ability at the individual level.

In sum, the questionnaire results of Study 2 replicated those in Study 1, providing further evidence that questionnaire measures of empathic ability support women's superiority over men. The EEG results revealed robust sex/gender differences in both phase-locked and non–phase-locked neural responses to face stimuli, which occurred in a wide time window from 70 to 550 ms after stimulus onset. The ERP results also showed robust evidence for enhanced phase-locked and non–phase-locked neural responses that were specific to painful expressions, replicating the findings of previous research (e.g. Sheng and Han, 2012; Sheng *et al.*, 2016; Zhou and Han, 2021). Nevertheless, the analyses of neither phase-locked nor non–phase-locked empathic neural responses to painful (*vs* neutral) expressions found evidence for reliable sex/gender differences. The results of Bayes factor analyses are consistent with the null hypothesis of sex/gender differences in empathic neural responses. Taken together, the results in Study 2 revealed inconsistent results in the questionnaire and EEG estimations of sex/gender differences in empathic ability.

Study 3: Effects of social expectations on questionnaire measures of empathic ability

A key question arising from the results in Studies 1 and 2 is why questionnaires but not EEG measures suggest sex/gender differences in empathic ability. Since questionnaire measures are susceptible to influences of social contexts or social desirability

(Heine *et al.*, 2002; Krumpal, 2013), it is likely that questionnaire measures of empathic ability are affected by social expectations of gender roles in sharing and caring for others' emotions. Therefore, in Study 3, we tested the hypothesis that sex/gender differences in questionnaire measures of empathic ability are influenced by social expectations. To manipulate social expectations of women or men's empathic ability, we created two essays to make salient women or men's empathic ability in the empathy-inducing priming condition and two essays in the control priming condition. An independent Chinese sample was recruited and randomly assigned to one of the four priming conditions. After reading a priming essay, each participant was asked to complete a distraction task. Thereafter, they completed the IRI questionnaire to test whether priming social expectations of empathy influence questionnaire estimation of sex/gender difference in empathic ability.

Methods
Participants
The sample size in Study 3 was pre-determined using G*Power estimation (Faul *et al.*, 2009). An average effect size for the sex/gender differences in IRI scores was $d = 0.38$ ($f = 0.19$) in Studies 1 and 2. This effect size was used for a power analysis. A sample size of 294 participants was required to obtain a small effect size of 0.19 for sex/gender difference with an error probability of 0.05 and a power of 0.90 for a two-way interaction of ANOVA. The power analysis of the two-way interaction was conducted due to a lack of prior data and effect size estimates. Thus, we recruited 328 Chinese college students in Study 3 (151 men, mean age $\pm$ s.d. $= 22.79 \pm 2.7$ years; 177 women, mean age $\pm$ s.d. $= 21.94 \pm 3.2$ years) via an online survey. All participants provided written informed consent after the experimental procedure had been fully explained and were informed of their right to withdraw at any time during the study. All participants were paid for their participation.

Stimuli and procedure
We created four essays to prime social expectations of empathic ability. Two empathy-inducing essays consist of fictional information that findings of scientific research show evidence that women (or men) are good at sharing and caring for the feelings of others. The two empathy-inducing essays were designed to relate specifically to empathic concern. Two control essays consist of fictional information that scientific research shows evidence that women (or men) are independent (see Supplementary materials for the four essays). The two control essays were designed to control the language and length of the essays without any contents about sharing and caring for the feelings of others. After the participants had filled out a form to report demographic information, they were asked to read one of the four essays to prime social expectations of empathy in men or in women. All participants were randomly assigned to one of the priming groups. Female participants read the essays related to women, and male participants read the essays related to men. Thereafter, the participants had to answer a question as a manipulation check of their understanding of the priming essays (44 participants do not pass this manipulation check, leaving 328 participants in total for further analysis: Empathy-inducing priming group: 76 men and 87 women; Control priming group: 75 men and 90 women). After the priming procedure, the participants completed a distraction task to solve three logical problems (e.g. finding a rule to present pictures sequentially). The distraction task served to remove the materials in the priming essays from working memory
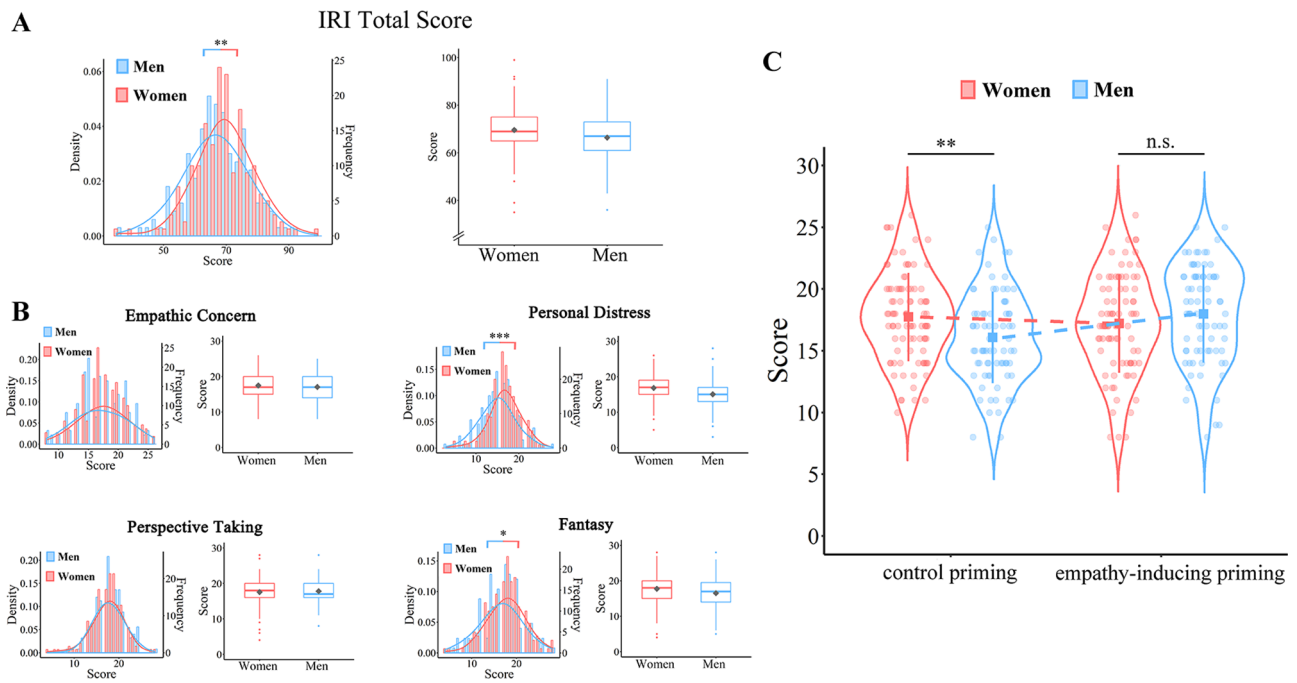
Fig. 6. Results of priming effects on questionnaire estimation of sex/gender differences in empathic ability in Study 3. (A) The density distribution and mean of the total IRI score. (B) The illustration of the density distribution and mean scores of each IRI subscale. The left panel shows the density (the left y-axis) and frequency (the right y-axis). Kernel density estimation was conducted to assess the probability density function of the questionnaire measures. The right panel illustrates bar charts of the total IRI score. The lower and upper hinges of the boxes correspond to the 25th and 75th percentiles, respectively. Each dot represents an outlier. The horizontal line inside each box shows the median. The lower and higher whiskers represent the lowest and highest observed values excluding outliers, respectively. The gray diamond in the middle represents the mean of the data. (C). The illustration of the priming effect on questionnaire estimations of sex/gender differences in empathic ability. Shown are the mean scores of the Empathic Concern subscales of IRI. Squares in the middle represent the mean of each group. Each dot represents one participant. The lower and higher whiskers represent the 1 s.d. away from the mean of each group, respectively. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. ns = not significant.

so as to reduce the direct effects of priming languages on subsequent questionnaire measures to a minimum degree. Finally, the participants were instructed to complete the IRI (Davis, 1983a) and self-construal questionnaire (Singelis, 1994) in a random order. The self-construal questionnaire was used to control for empathy-unrelated priming effects.

Results and discussion

We first tested sex/gender differences in questionnaire measures of empathic ability without considering the priming effect by conducting independent two-sample *t*-tests of IRI rating scores. The results showed that women (*vs* men) reported significantly higher IRI total scores ($t(326) = 3.080$, $P = 0.005$, FDR corrected, Cohen's $d = 0.341$, 95% CI = [0.122–0.560], Figure 6A). Women (*vs* men) also scored significantly higher on the subscales including Personal Distress ($t(326) = 4.251$, $P < 0.001$, FDR corrected, Cohen's $d = 0.471$, 95% CI = [0.250–0.691]) and Fantasy ($t(326) = 2.626$, $P = 0.015$, FDR corrected, Cohen's $d = 0.291$, 95% CI = [0.072–0.509]) but not on the subscales including Empathic Concern ($t(326) = 1.018$, $P = 0.386$, FDR corrected, Cohen's $d = 0.113$, 95% CI = [−0.105–0.330]) and Perspective Taking subscale ($t(326) = −0.667$, $P = 0.505$, Cohen's $d = −0.074$, 95% CI = [−0.291–0.143]). These results replicated the findings of sex/gender difference in questionnaire measures of empathic ability in Studies 1 and 2 except that the sex/gender difference in scores of Empathic Concern showed a similar trend but did not reach significance, possibly due to the small sample size in Study 3.

Because the priming essays focused on the ability of sharing and caring for others' feelings, we expected that priming

essays might reduce sex/gender differences in rating scores of the Empathic Concern subscale of IRI. We conducted a two-way ANOVA of the scores with Priming (empathy inducing *vs* control priming) and Gender (women *vs* men) as two between-subjects variables. The results showed a significant interaction between Priming and Gender ($F(1324) = 8.323$, $P = 0.004$, $\eta_p^2 = 0.025$, 90% CI = [0.005–0.059], Figure 6C), suggesting distinct patterns of sex/gender differences in the rating scores of the Empathic Concern subscale in the two priming conditions. Simple effect comparisons of the rating scores further revealed that women compared to men reported larger empathic concern scores in the control priming condition ($F(1324) = 7.718$, $P = 0.006$, $\eta_p^2 = 0.023$, 90% CI = [0.004–0.057]) but not in the empathy-inducing condition ($F(1324) = 1.703$, $P = 0.193$, $\eta_p^2 = 0.005$, 90% CI = [0.000–0.026]). In addition, the comparisons between the rating scores in the two priming conditions showed significantly higher scores in the empathy-inducing condition than in the control priming condition in men ($F(1324) = 9.473$, $P = 0.002$, $\eta_p^2 = 0.028$, 90% CI = [0.006–0.064]) but not in women ($F(1324) = 0.846$, $P = 0.358$, $\eta_p^2 = 0.003$, 90% CI = [0.000–0.020]). These results suggest that the empathy-inducing compared to the control priming reduced sex/gender difference in the empathic concern score by increasing men's self-report of their empathic concern.

We also tested the priming effects on the rating scores of other subscales. The results showed significant main effects of gender on rating scores of both the Personal Distress ($F(1324) = 17.918$, $P < 0.001$, $\eta_p^2 = 0.052$, 90% CI = [0.020–0.096]) and Fantasy ($F(1324) = 6.847$, $P = 0.009$, $\eta_p^2 = 0.021$, 90% CI = [0.003–0.053]) subscale but not of the Perspective Taking subscale ($F(1324) = 0.442$, $P = 0.507$, $\eta_p^2 = 0.001$, 90% CI = [0.000–0.016]), suggesting larger

rating scores of Personal Distress and Fantasy in women than in men. However, there was neither a significant main effect of priming nor its interaction with gender on rating scores of any of these three subscales (*Ps* > 0.5), providing no evidence for influences of the empathy inducing *vs* control priming.

Similar analyses of the rating scores of independence and interdependence failed to show reliable sex/gender differences in either the empathy-inducing condition or the control priming condition (see Supplementary materials for details). These results showed evidence that activation of social expectations of both women and men's ability of sharing and caring others' feelings diminishes the sex/gender difference in questionnaire estimation of empathic ability. These results are consistent with the hypothesis that sex/gender differences in questionnaire measures of empathic ability are influenced by social expectations and implicate possible contributions of default social desirability of women's better empathic ability to the observed sex/gender differences in empathic ability in the previous questionnaire studies.

General discussion

In three studies, we investigated sex/gender differences in empathic ability by collecting questionnaire (IRI) and EEG (both phase-locked and non–phase-locked signals) measures of empathic ability. Our questionnaire measures showed higher total scores of IRI and scores of some of the subscales in women than in men. The sex/gender difference in IRI estimation of empathic ability was repeatedly observed in three independent samples with different sample sizes. These results were consistent with previous findings based on either IRI (Davis, 1980; Eisenberg and Lennon, 1983; Thompson and Voyer, 2014) or the Empathy Quotient (Baron-Cohen and Wheelwright, 2004; Lawrence *et al.*, 2004; Greenberg *et al.*, 2018). Thus, subjective (questionnaire) estimations of empathic ability lend support to the notion of women's superiority over men in empathy, and this observation appears to be independent of questionnaires used and cultural samples tested.

However, our EEG estimations of empathic ability based on a large sample of EEG measures support a null hypothesis of sex/gender difference in empathic ability. Our previous EEG study of a small subject sample failed to show a significant difference in ERP amplitudes in response to perceived painful (*vs* non-painful) stimulations to others' hands between male and female participants (Han *et al.*, 2008). Our EEG results in the current study with a larger sample further showed evidence for comparable empathic neural responses in both phase-locked and non-phase-locked signals within 600 ms after stimulus onset in women and men. A null hypothesis of sex/gender difference in empathic neural response was supported by the results of Bayesian analyses of our EEG data. The lack of sex/gender difference in empathic neural responses cannot be due to the fact that EEG signals were insensitive to our experimental manipulations. Our EEG results showed that both phase-locked (i.e. P2 amplitudes) and non–phase-locked (4–6 Hz oscillations) neural responses were modulated by painful (*vs* neutral) expressions as early as 140 ms after face onset similar to previous EEG findings (Mu *et al.*, 2008; Sheng and Han, 2012; Sheng *et al.*, 2016). The amplitudes of long-latency ERP components (i.e. N2 and LPP) were also significantly modulated by painful (*vs* neutral) expressions, consistent with previous ERP findings (Sessa *et al.*, 2014; Sessa and Meconi, 2015; Cui *et al.*, 2016, 2017). In addition, both phase-locked (i.e. N1 and N170 amplitudes) and non–phase-locked (15–27 Hz oscillations)

brain activities demonstrated enhanced neural responses to face stimuli in women than in men. These results indicate that our EEG measures were sensitive enough to reveal empathic neural responses and sex/gender differences in brain activities in response to face stimuli. Nevertheless, comparisons of empathic neural responses in women and men failed to reveal any reliable sex/gender difference in phase-locked ERP amplitudes or powers of non–phase-locked oscillations that were sensitive to others' painful feelings. Importantly, the null hypothesis regarding sex/gender difference in these empathic neural responses was further supported by the results of our Bayes analyses. Our correlation analyses across individuals did not find any evidence for associations between questionnaire and EEG measures of empathic ability. These results provide further evidence for a dissociation between the subjective and objective estimation of empathic ability.

Given the contradictory results of questionnaire and brain imaging estimations of sex/gender differences in empathic ability, we tested a mechanical interpretation of the observed sex/gender difference in questionnaire measures of empathic ability. We empirically tested an account that the sex/gender difference in empathic ability shown in questionnaire measures is caused by social expectations of gender role in caring for others' feelings. Our results in Study 3 showed that, while questionnaire measures in a control condition showed higher rating scores in women than in men, this sex/gender difference was eliminated by priming social expectations that both women and men are good at sharing and caring for the feelings of others. The results of the priming manipulations suggest that priming social expectations of men's ability to share and care about others' feelings tended to increase men's rating scores of IRI affective subscales, resulting in the absence of sex/gender differences in empathic ability estimated by questionnaires. The priming effects observed here are consistent with previous brain imaging findings of greater impact of social contextual cues on empathetic brain responses in men than in women (Ickes *et al.*, 2000; Singer, 2006)) Our findings suggest a causal relationship between social expectations and rating scores of questionnaire measures of empathic ability.

It has been realized by researchers that the findings of sex/gender differences in cognition and behavior allow inferences of the average difference between women and men rather than gender-based predictions about an individual's psychological capacity (e.g. Greenberg *et al.*, 2018). The findings of the inconsistent questionnaire and EEG results regarding sex/gender differences in empathy in the current work further indicate that researchers should be careful to apply the conclusion of sex/gender differences obtained from self-reported measurements to evaluate an individual's empathic ability. Integrating both subjective and objective estimations may provide useful information about an individual's psychological capacity such as the empathic ability.

Previous studies investigated the relationships between empathy and hormonal states in women but showed inconsistent results. For instance, it was found that, while women with low estradiol and progesterone levels (days 2–5 of menstrual cycle) showed higher accuracies of recognition of facial expressions than those with high estradiol and progesterone (days 18–25 of menstrual cycle), questionnaire measures of empathy did not differ significantly between the two groups (Derntl *et al.*, 2013). Another work testing emotional memory as empathy-related measures did not find a significant difference in memory performances between women in the mid-cycle and the later cycle

(Gamsakhurdashvili *et al.*, 2021). We also examined whether our questionnaire and EEG measures of empathy varied significantly with days of menstrual cycle across female participants but failed to find a significant effect (see Figure S2). Thus, similar to previous studies, our work provided no evidence for systematical variation of empathic ability due to hormonal state changes at different phases of women's menstrual cycle, which seems not a critical factor that accounts for our observed sex/gender differences in questionnaire measures of empathy.

The results of our work raised new issues that should be addressed in future research. First, previous work showed that empathic neural responses revealed in EEG signals to both painful expressions and painful stimuli applied to others' body parts are associated with self-report of sharing others' feelings (Fan and Han, 2008; Sheng and Han, 2012). The current work only recorded EEG signals to painful expressions. Future work may collect EEG signals to painful stimuli applied to others' body parts from a large sample to examine sex/gender differences in empathic neural responses as a different type of objective measure of empathic ability. Second, previous research examining interindividual variability in IRI scores showed that rating scores of different IRI subscales are possibly associated with different brain structures. For example, individual differences in affective empathic abilities oriented toward others were negatively correlated with gray matter volume in the inferior frontal gyrus and anterior cingulate, whereas gray matter volume of the anterior cingulate predicts better cognitive perspective-taking abilities (Banissy *et al.*, 2012). A key question arising from these results is whether women and men employ the same neural network when conducting questionnaire measures of empathic ability. This can be addressed by collecting functional brain imaging data when women and men perform questionnaire measures of empathic ability. Such brain imaging data would help to clarify and provide a neuroscience account of the sex/gender differences observed in questionnaire measures of psychological traits. Finally, previous fMRI studies that examined sex/gender differences in various brain structures and functional activities showed that, while males' brains are larger than females' from birth, stabilizing ~11% in adults, task-based fMRI has failed to find reproducible activation differences between men and women in tasks engaging verbal, spatial or emotion processing (Eliot *et al.*, 2021). While these fMRI findings challenge the concept of sexual dimorphism of the human brain, our ERP results showed evidence for larger N1/P1/N170/LPP amplitudes in responses to faces in women (*vs* men). These EEG findings raise the question of whether functional activities in the brain regions underlying multiple stages of face processing are stronger in women than in men or whether the observed sex/gender differences in ERP amplitudes arise from the sex/gender differences in structures of the brain and scalp that may influence how well neural signals are recorded at electrodes on the scalp.

There were a few limitations of the current work. First, our work did not rule out the impacts of other mental processes on sex/gender differences in the subjective estimation of empathic ability. For example, it has been shown that sex/gender differences in self-reported empathic ability increased when the motivation for empathy was raised (Löffler and Greitemeyer, 2021). In our Study 2, women, compared to men, showed better performance efficiencies (e.g. faster responses and higher response accuracies) when responding to painful expressions during EEG recording. This result might be due to women's greater motivation for good performance, which might also contribute to

the observed sex/gender difference in subjective (but not objective) estimation of empathic ability. Recent research has reported empirical findings that suggest sex/gender differences in strategies of emotion regulation and underlying brain activities (e.g. ; Goubet and Chrysikou, 2019), which may also influence how women and men report rating scores regarding questions of own empathic ability when being tested with questionnaires. Second, the task of judgments of painful *vs* neutral expressions puts a strong focus on top-down driven responses to others' emotional states. The bottom-up or automatic processes of empathy (Cuff, *et al.*, 2016) may not be uncovered well in this task. Third, as the priming task used in Study 3 focused on empathic concern, the results left an open issue regarding whether sex/gender differences in other aspects of empathy are similarly sensitive to social desirability. Fourth, our work tested only college students. Given previous findings of age differences in self-reported measures of empathy that, though, showed mixed results (e.g. Grühn *et al.*, 2008; Khanjani *et al.*, 2015), future work should examine sex/gender differences in empathy in other age groups. It is necessary to clarify whether the dissociation between questionnaire and EEG measures regarding sex/gender differences in empathic ability also exists in other samples such as adolescents and elderly adults. Finally, although our EEG measures were sensitive to perceived painful expressions in multiple time windows, the EEG results were unable to disentangle empathic neural responses in specific brain regions due to their low spatial resolution. Previous fMRI studies suggested greater empathic neural responses in women in some brain regions but in men in other brain regions (Schulte-Rüther *et al.*, 2008; Derntl *et al.*, 2010; Christov-Moore and Iacoboni, 2019). Although the conclusions about sex/gender differences in empathic neural responses in previous fMRI studies are limited by the small testing samples, the fMRI findings suggest that women and men might take distinct mental strategies supported by different brain regions during empathy for others' emotional states. Large sample fMRI studies are required to clarify this in future research.

In conclusion, by collecting questionnaires and EEG measures of empathic ability, we showed evidence that subjective and objective measures gave different conclusions regarding sex/gender differences in empathic ability in young adults. Questionnaire measures of empathic ability suggest women's superiority, whereas EEG measures of empathic ability support a null hypothesis regarding sex/gender differences in empathy. In addition, we empirically tested and provided evidence that social expectations contribute to the observed sex/gender differences in questionnaire measures of empathic ability. Our results indicate that whether women are more empathetic and caring than men remains an open question. Future research should be careful when making inferences about social and clinical implications of the sex/gender difference in empathic ability suggested by questionnaire measures.

Supplementary data

Supplementary data are available at *SCAN* online.

Data availability

All data and code in this and the following studies are available at https://osf.io/srhke/ according to Institutional Review Board restrictions regarding participant privacy/consent.

Conflict of interest

The authors declared that they had no conflict of interest with respect to their authorship or the publication of this article.

References

Avenanti, A., Bueti, D., Galati, G., Aglioti, S.M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience*, **8**, 955–60.

Baez, S., Flichtentrei, D., Prats, M., *et al.* (2017). Men, women…who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLoS One*, **12**(6), e0179336.

Banissy, M.J., Kanai, R., Walsh, V., Rees, G. (2012). Inter-individual differences in empathy are reflected in human brain structure. *Neuroimage*, **62**(3), 2034–9.

Baron-Cohen, S., Knickmeyer Rebecca, C., Belmonte Matthew, K. (2005). Sex differences in the brain: implications for explaining autism. *Science*, **310**(5749), 819–23.

Baron-Cohen, S., Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, **34**(2), 163–75.

Batson, C.D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Batson, C.D. (2011). *Altruism in Humans*. Oxford: Oxford University Press.

Bekkali, S., Youssef, G.J., Donaldson, P.H., Albein-Urios, N., Hyde, C., Enticott, P.G. (2021). Is the putative mirror neuron system associated with empathy? A systematic review and meta-analysis. *Neuropsychology Review*, **31**(1), 14–57.

Berman, P.W. (1980). Are women more responsive than men to the young? A review of developmental and situational variables. *Psychological Bulletin*, **88**(3), 668–95.

Cheng, Y., Yang, C.Y., Lin, C.P., Lee, P.L., Decety, J. (2008). The perception of pain in others suppresses somatosensory oscillations: a magnetoencephalography study. *NeuroImage*, **40**(4), 1833–40.

Christov-Moore, L., Iacoboni, M. (2019). Sex differences in somatomotor representations of others' pain: a permutation-based analysis. *Brain Structure and Function*, **224**(2), 937–47.

Christov-Moore, L., Simpson, E.A., Coudé, G., Grigaityte, K., Iacoboni, M., Ferrari, P.F. (2014). Empathy: gender effects in brain and behavior. *Neuroscience and Biobehavioral Reviews*, **46**, 604–27.

Chrysikou, E. G., Thompson, W. J. (2016). Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model. *Assessment*, **23**(6), 769–77.

Cuff, B. M. P., Brown, S. J., Taylor, L., Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review*, **8**(2), 144–53.

Cui, F., Abdelgabar, A.-R., Keysers, C., Gazzola, V. (2015). Responsibility modulates pain-matrix activation elicited by the expressions of others in pain. *NeuroImage*, **114**(7), 371–8.

Cui, F., Zhu, X., Duan, F., Luo, Y. (2016). Instructions of cooperation and competition influence the neural responses to others' pain: an ERP study. *Social Neuroscience*, **11**(3), 289–96.

Cui, F., Zhu, X., Luo, Y. (2017). Social contexts modulate neural responses in the processing of others' pain: an event-related potential study. *Cognitive, Affective & Behavioral Neuroscience*, **17**(4), 850–7.

Davis, M.H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, **10**, 85.

Davis, M.H. (1983a). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, **44**(1), 113–26.

Davis, M.H. (1983b). Empathic concern and the muscular dystrophy telethon: empathy as a multidimensional construct. *Personality & Social Psychology Bulletin*, **9**(2), 223–9.

Decety, J., Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, **3**(2), 71–100.

Decety, J., Michalska, K.J., Akitsuki, Y. (2008). Who caused the pain? An fMRI investigation of empathy and intentionality in children. *Neuropsychologia*, **46**(11), 2607–14.

Derntl, B., Finkelmeyer, A., Eickhoff, S., *et al.* (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology*, **35**(1), 67–82.

Derntl, B., Hack, R.L., Kryspin-Exner, I., Habel, U. (2013). Association of menstrual cycle phase with the core components of empathy. *Hormones and Behavior*, **63**(1), 97–104.

Dienes, Z. (2011). Bayesian versus orthodox statistics: which side are you on? *Perspectives on Psychological Science*, **6**(3), 274–90.

Ding, R., Ren, J., Li, S., Zhu, X., Zhang, K., Luo, W. (2020). Domain-general and domain-preferential neural correlates underlying empathy towards physical pain, emotional situation and emotional faces: an ALE meta-analysis. *Neuropsychologia*, **137**, 107286.

Domes, G., Schulze, L., Böttger, M., *et al.* (2010). The neural correlates of sex differences in emotional reactivity and emotion regulation. *Human Brain Mapping*, **31**(5), 758–69.

Dorrough, A.R., Glöckner, A. (2019). A cross-national analysis of sex differences in prisoner's dilemma games. *British Journal of Social Psychology*, **58**(1), 225–40.

Efron, B., Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*, 1st edn. New York: Chapman and Hall/CRC.

Eisenberg, N., Fabes, R.A. (1990). Empathy: conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, **14**(2), 131–49.

Eisenberg, N., Lennon, R. (1983). Sex differences in empathy and related capacities [Article]. *Psychological Bulletin*, **94**(1), 100–31.

Eliot, L., Ahmed, A., Khan, H., Patel, J. (2021). Dump the "dimorphism": comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience and Biobehavioral Reviews*, **125**(6), 667–97.

Fabi, S., Leuthold, H. (2017). Empathy for pain influences perceptual and motor processing: evidence from response force, ERPs, and EEG oscillations. *Social Neuroscience*, **12**(6), 701–16.

Fallon, N., Roberts, C., Stancak, A. (2020). Shared and distinct functional networks for empathy and pain processing: a systematic review and meta-analysis of fMRI studies. *Social Cognitive and Affective Neuroscience*, **15**(7), 709–23.

Fan, Y., Duncan, N.W., de Greck, M., Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews*, **35**(3), 903–11.

Fan, Y., Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia*, **46**(1), 160–73.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, **41**(4), 1149–60.

Gamsakhurdashvili, D., Antov, M.I., Stockhorst, U. (2021). Sex-hormone status and emotional processing in healthy women. *Psychoneuroendocrinology*, **130**, 105258.

Gilet, A.-L., Mella, N., Studer, J., Grühn, D., Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, **45**(1), 42–8.

Goubet, K.E., Chrysikou, E.G. (2019). Emotion regulation flexibility: gender differences in context sensitivity and repertoire. *Frontiers in Psychology*, **10**, 935.

Greenberg, D.M., Warrier, V., Allison, C., Baron-Cohen, S. (2018). Testing the Empathizing-Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences*, **115**(48), 12152–7.

Grossmann, A., Kronland-Martinet, R., Morlet, J. (1990). *Reading and Understanding Continuous Wavelet Transforms*. Berlin, Heidelberg: Wavelets.

Grühn, D., Rebucal, K., Diehl, M., Lumley, M., Labouvie-Vief, G. (2008). Empathy across the adult lifespan: longitudinal and experience-sampling findings. *Emotion*, **8**(6), 753–65.

Gu, X., Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *Neuroimage*, **36**(1), 256–67.

Han, S., Fan, Y., Mao, L. (2008). Gender difference in empathy for pain: an electrophysiological investigation. *Brain Research*, **1196**(2), 85–93.

Heine, S.J., Lehman, D.R., Peng, K., Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, **82**(6), 903–18.

Hoffman, M.L. (2008). *Empathy and Prosocial Behavior*. New York: The Guilford Press.

Huang, X., Li, W., Sun, B., Chen, H., Davis, M.H. (2012). The validation of the interpersonal reactivity index for Chinese teachers from primary and middle schools. *Journal of Psychoeducational Assessment*, **30**(2), 194–204.

Ickes, W., Gesn, P.R., Graham, T. (2000). Gender differences in empathic accuracy: differential ability or differential motivation? *Personal Relationships*, **7**(1), 95–109.

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine*, **2**(8), e124.

Jackson, P.L., Decety, J. (2004). Motor cognition: a new paradigm to study self-other interactions *Current Opinion in Neurobiology*, **14**, 259–63.

Jauniaux, J., Khatibi, A., Rainville, P., Jackson, P.L. (2019). A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience*, **14**(8), 789–813.

Joyal, C.C., Neveu, S., Boukhalfi, T., Jackson, P.L., Renaud, P. (2018). Suppression of Sensorimotor Alpha Power Associated With Pain Expressed by an Avatar: A Preliminary EEG Study. *Frontiers in Human Neuroscience*, **12**, 273.

Khanjani, Z., Mosanezhad Jeddi, E., Hekmati, I., *et al.* (2015). Comparison of cognitive empathy, emotional empathy, and social

functioning in different age groups. *Australian Psychologist*, **50**(1), 80–5.

Klein, K.J.K., Hodges, S.D. (2001). Gender differences, motivation, and empathic accuracy: when it pays to understand. *Personality & Social Psychology Bulletin*, **27**(6), 720–30.

Kronland-Martinet, R., Morlet, J., Grossmann, A. (1987). Analysis of sound patterns through Wavelet transforms. *International Journal of Pattern Recognition and Artificial Intelligence*, **01**(02), 273–302.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, **47**(4), 2025–47.

Lamm, C., Decety, J., Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage*, **54**(3), 2492–502.

Lamm, C., Nusbaum, H.C., Meltzoff, A.N., Decety, J., Warrant, E. (2007). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS One*, **2**(12), e1292.

Lamm, C., Rütgen, M., Wagner, I.C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters*, **693**, 49–53.

Lawrence, E.J., Shaw, P., Baker, D., Baron-Cohen, S., David, A.S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine*, **34**(5), 911–9.

Levy, J., Goldstein, A., Influs, M., Masalha, S., Zagoory-Sharon, O., Feldman, R. (2016). Adolescents growing up amidst intractable conflict attenuate brain response to pain of outgroup. *Proceedings of the National Academy of Sciences*, **113**, 13696–701.

Li, X., Liu, Y., Ye, Q., Lu, X., Peng, W. (2020). The linkage between first-hand pain sensitivity and empathy for others' pain: attention matters. *Human Brain Mapping*, **41**(17), 4815–28.

Liu, S., Seidlitz, J., Blumenthal, J.D., Clasen, L.S., Raznahan, A. (2020). Integrative structural, functional, and transcriptomic analyses of sex-biased brain organization in humans. *Proceedings of the National Academy of Sciences*, **117**(31), 18788–98.

Löffler, C.S., Greitemeyer, T. (2023). Are women the more empathetic gender? The effects of gender role expectations. *Current Psychology* **42**, 220–31.

Luders, E., Toga, A.W. (2010). Sex differences in brain anatomy. *Progress in Brain Research*, **186**, 3–12.

Luo, S., Li, B., Ma, Y., Zhang, W., Rao, Y., Han, S. (2015). Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. *Neuroimage*, **110**(4), 22–31.

Masten, C.L., Morelli, S.A., Eisenberger, N.I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *Neuroimage*, **55**(1), 381–8.

Mathur, V.A., Harada, T., Lipke, T., Chiao, J.Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, **51**(4), 1468–75.

Mehrabian, A., Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, **40**(4), 525–43.

Mestre, M., Samper, P., Frías, M., Tur, A. (2009). Are Women More Empathetic than Men? A Longitudinal Study in Adolescence. *The Spanish Journal of Psychology*, **12**(1), 76–83.

Morey, R.D., Rouder, J.N. (2018). BayesFactor: computation of Bayes factors for common designs. R package version. In (Version R package version 0.9.12-4.2). Available: https://CRAN.R-project.org/package=BayesFactor [05 July 2022].

Motoyama, Y., Ogata, K., Hoka, S., Tobimatsu, S. (2017). Frequency-dependent changes in sensorimotor and pain affective systems induced by empathy for pain. *Journal of Pain Research*, **10**, 1317–26.

Mu, Y., Fan, Y., Mao, L., Han, S. (2008). Event-related theta and alpha oscillations mediate empathy for pain. *Brain Research*, **1234**(10), 128–36.

Murphy, B.A., Lilienfeld, S.O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment*, **31**(8), 1062–72.

National Institutes of Health. (2015). Consideration of sex as a biological variable in NIH- funded Research. Available: https://orwh.od.nih.gov/sites/orwh/files/docs/NOT-OD-15-102_Guidance.pdf [09 June 2015].

Olsson, M.I.T., Froehlich, L., Dorrough, A.R., Martiny, S.E. (2021). The hers and his of prosociality across 10 countries. *British Journal of Social Psychology*, **60**(4), 1330–49.

Palmieri, A., Meconi, F., Vallesi, A., *et al.* (2020). Enhanced neural empathic responses in patients with spino-bulbar muscular atrophy: an electrophysiological study. *Brain Sciences*, **11**(1), 16.

Perry, A., Bentin, S., Bartal, I.BA. *et al.* (2010). "Feeling" the pain of those who are different from us: Modulation of EEG in the mu/alpha range. *Cognitive, Affective, & Behavioral. Neuroscience*, **10**, 493–504.

Pfurtscheller, G., Lopes da Silva, F.H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, **110**(11), 1842–57.

Preis, M.A., Kroener-Herwig, B. (2012). Empathy for pain: the effects of prior experience and sex. *European Journal of Pain*, **16**(9), 1311–9.

Riečanský, I., Lamm, C. (2019). The role of sensorimotor processes in pain empathy. *Brain Topography*, **32**(6), 965–76.

Riečanský, I., Lengersdorff, L.L., Pfabigan, D.M., Lamm, C. (2020). Increasing self-other bodily overlap increases sensorimotor resonance to others' pain. *Cognitive, Affective & Behavioral Neuroscience*, **20**(1), 19–33.

Rueckert, L., Branch, B., Doan, T. (2011). Are gender differences in empathy due to differences in emotional reactivity? *Psychology*, **2**(6), 574–8.

Ruigrok, A.N.V., Salimi-Khorshidi, G., Lai, M.-C., *et al.* (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and Biobehavioral Reviews*, **39**(2), 34–50.

Schulte-Rüther, M., Markowitsch, H.J., Shah, N.J., Fink, G.R., Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, **42**(1), 393–403.

Sessa, P., Meconi, F. (2015). Perceived trustworthiness shapes neural empathic responses toward others' pain. *Neuropsychologia*, **79**, 97–105.

Sessa, P., Meconi, F., Castelli, L., Dell'Acqua, R. (2014). Taking one's time in feeling other-race pain: an event-related potential investigation on the time-course of cross-racial empathy. *Social Cognitive and Affective Neuroscience*, **9**(4), 454–63.

Sheng, F., Han, S. (2012). Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *NeuroImage*, **61**(4), 786–97.

Sheng, F., Han, X., Han, S. (2016). Dissociated neural representations of pain expressions of different races. *Cerebral Cortex*, **26**(3), 1221–33.

Sheng, F., Liu, Q., Li, H., Fang, F., Han, S. (2014). Task modulations of racial bias in neural responses to others' suffering. *NeuroImage*, **88**(3), 263–70.

Singelis, T.M. (1994). The measurement of independent and interdependent self-construals. *Personality & Social Psychology Bulletin*, **20**(5), 580–91.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, **30**(6), 855–63.

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R.J., Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, **303**(5661), 1157–62.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, **439**(7075), 466–9.

Steiger, J.H. (2004). Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, **9**(2), 164–82.

Sun, Y., Gao, X., Han, S. (2010). Sex differences in face gender recognition: an event-related potential study. *Brain Research*, **1327**(4), 69–76.

Thompson, A.E., Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: a meta-analysis. *Cognition & Emotion*, **28**(7), 1164–95.

Whitmarsh, S., Nieuwenhuis, I.L., Barendregt, H.P., Jensen, O. (2011). Sensorimotor Alpha Activity is Modulated in Response to the Observation of Pain in Others. *Frontiers in Human. Neuroscience*, **5**, 91.

Xu, X., Zuo, X., Wang, X., Han, S. (2009). Do You Feel My Pain? Racial Group Membership Modulates Empathic Neural Responses. *Journal of Neuroscience*, **29**, 8525–9.

Yang, C.-Y., Decety, J., Lee, S., Chen, C., Cheng, Y. (2009). Gender differences in the mu rhythm during empathy for pain: an electroencephalographic study. *Brain Research*, **1251**(1), 176–84.

Yang, H., Kang, SJ. 2020 Exploring the Korean adolescent empathy using the Interpersonal Reactivity Index (IRI). *Asia Pacific Education Review*, **21**, 339–49.

Zhao, Q., Neumann, D. L., Cao, X., Baron-Cohen, S., Sun, X., Cao, Y., Yan, C., Wang, Y., Shao, L., Shum, D. H. K. (2018). Validation of the Empathy Quotient in Mainland China. *Journal of Personality Assessment*, **100**(3), 333–42.

Zhou, Y., Han, S. (2021). Neural dynamics of pain expression processing: alpha-band synchronization to same-race pain but desynchronization to other-race pain. *Neuroimage*, **224**(1), 117400.