NeuroImage 209 (2020) 116488

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Guilty by association: How group-based (collective) guilt arises in the brain

Zhiai Li[a,b,1], Hongbo Yu[b,c,1], Yongdi Zhou[a], Tobias Kalenscher[d], Xiaolin Zhou[b,e,f,g,*]

[a] School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China
[b] School of Psychological and Cognitive Sciences, Peking University, Beijing, 100871, China
[c] Department of Psychology, Yale University, New Haven, CT, 06520, USA
[d] Comparative Psychology, Heinrich Heine University, Düsseldorf, Germany
[e] Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China
[f] Institute of Psychological and Brain Sciences, Zhejiang Normal University, Zhejiang, 321004, China
[g] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China

ARTICLE INFO

*Keywords:*
Group-based guilt
Minimal group paradigm
Perceived responsibility
fMRI
Cingulate cortex

ABSTRACT

People do not only feel guilty for transgressions that they are causally responsible for (i.e., personal guilt); they also feel guilty for transgressions committed by those they identify as in-group members (i.e., collective or group-based guilt). Decades of research using scenario-based imagination tasks and self-reported measures has shown that when reminded of transgressions committed by in-group members, people express guilt and are willing to make amends, even when they are not causally responsible for the transgressions. However, it remains elusive whether people genuinely experience guilt or simply display remorseful gestures deemed appropriate in those contexts. To resolve this puzzle, it is critical to closely examine the neurocognitive basis of group-based guilt and its relationship with personal guilt, a goal that self-reported measures alone cannot satisfactorily achieve. Here, we combined functional MRI with an interaction-based minimal group paradigm in which participants either directly caused harm to a group of victims (i.e., personal guilt), or observed in-group members cause harm to the victims (i.e., group-based guilt). In three experiments ($N = 90$), we demonstrated and replicated that the perceived responsibility one shared with in-group members in transgression predicted both behavioral and neural manifestations of group-based guilt. Multivariate pattern analysis (MVPA) of the functional MRI data showed that group-based guilt recruited patterns of neural responses in anterior middle cingulate cortex that resembled personal guilt. These results have broadened our understanding of how group membership is integrated into the neurocognitive processes underlying social emotions.

## 1. Introduction

Guilt is viewed as “the emotion most essential to the development of conscience and moral behavior” in field of psychology (Izard, 1991) and as the “internalized voice of moral authority” in field of philosophy (Griswold, 2007). People feel guilty when they realize that they are responsible for an action or omission that violates moral norms or mutual expectations that they accept as binding (i.e., personal guilt) (Baumeister et al., 1994, 1995; Tangney and Dearing, 2003; Taylor, 1985). Guilt can also be encountered in inter-group interactions (Halperin and Schori-Eyal, 2019; Vollberg and Cikara, 2018): individuals may feel guilty for transgressions committed by members of social groups they identify as in-group, even when he/she is not directly responsible for these transgressions. However, the psychological and neural basis of group-based (collective) guilt and its relation to personal guilt are poorly understood.

As individuals rarely engage in social interactions without social identity or association (Mesquita et al., 2016; Tajfel and Turner, 1986), social emotions arising from such interactions are often tainted by group identity and inter-group appraisals (Mackie et al., 2008; Smith and Mackie, 2015). Well-known cases of group-based or collective social emotions have been widely debated and reflected upon theoretically (Smiley, 2017; Perron Tollefsen, 2003), and have received extensive empirical investigation in the past decades (Branscombe et al., 2004; Doosje et al., 1998; Ferguson and Branscombe, 2014; Wohl et al., 2006). In this line of research, the most frequently used method for inducing group-based guilt is scenario-based imagination or recall of historical events involving intergroup conflict (Brown et al., 2008; Doosje et al.,

[*] Corresponding author. School of Psychological and Cognitive Sciences, Peking University, 5 Yiheyuan Road, Beijing, 100871, China.
*E-mail address:* xz104@pku.edu.cn (X. Zhou).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2019.116488
Received 13 September 2019; Received in revised form 19 December 2019; Accepted 20 December 2019
Available online 27 December 2019

1998; McGarty et al., 2005). These studies demonstrated that group-based guilt results from the acceptance of in-group responsibility for transgressions (Castano and Giner-Sorolla, 2006; Čehajić-Clancy et al., 2011), could facilitate inter-group reconciliation (Allpress et al., 2014; Doosje et al., 2004; Halperin and Schori-Eyal, 2019; Lickel et al., 2011; Wohl et al., 2019), and reduce prejudice towards out-group (Amodio et al., 2007; Powell et al., 2005).

Although the scenario-based approach has consistently shown that self-reported guilt is elicited and group-based responsibility is perceived when participants are reminded of in-group misdeeds (Brown et al., 2008; Doosje et al., 1998; McGarty et al., 2005), the psychological nature of this pattern remains elusive: (1) Does the self-reported guilt reflect genuine feelings or does it merely reflect what the participants find morally appropriate to express? That is, do participants express guilt-like sentiments in the absence of genuine feelings of guilt, only to meet social expectations dictating the expression thereof? (2) How does the brain encode group-based guilt? Specifically, does group-based guilt share common neurocognitive processes with personal guilt?

According to the Intergroup Emotion Theory (IET; Mackie et al., 2008; Smith and Mackie, 2015), group-based emotions are similar to individual-level emotions in terms of their cognitive antecedents, phenomenological experience and action tendency (Rydell et al., 2008). In contrast, according to the Display Rules Hypothesis (Diefendorff and Richard, 2003; Matsumoto, 1993), it is sometimes socially desirable or even morally required to express certain emotions in specific contexts. Even when an individual does not genuinely experience the emotion, they will nevertheless display it to comply with social/moral norms. To distinguish these hypotheses, self-report of emotion alone is not enough. We therefore recorded participants' brain responses in guilt-eliciting social interactions (Koban et al., 2013; Yu et al., 2014). Importantly, in these social interaction tasks the participants were not required to report guilt; indeed, the term "guilt" was not even mentioned. Therefore, the participants were not incentivized to feel or express guilt in this context. Leveraging this paradigm, previous neuroimaging studies have consistently identified activations in cingulate cortex and insula as critical neural substrates underlying guilt-eliciting social interactions (Cui et al., 2015; Koban et al., 2013; Radke et al., 2011; Yu et al., 2014). Moreover, a meta-analysis reported in Yu et al. (2014) demonstrated that the anterior middle cingulate (aMCC) and anterior insula (AI) activations induced by guilt-eliciting social interactions are spatially separated from brain structures related to executive control, norm-compliance and emotion regulation, such as dorsal anterior cingulate cortex (dACC) and lateral prefrontal cortex (LPFC) (Buckholtz, 2015; Crockett et al., 2017; Goldin et al., 2008; McRae et al., 2010). The specific neural hypothesis we aimed to test in the current study is based on these observations, namely, if people genuinely experience guilt in an in-group transgression situation, we should observe neural activations resembling the patterns observed in the personal guilt situation (i.e., aMCC, AI). If, on the other hand, people merely feel obliged to express concerns in the in-group transgression situation, we should observe neural activations in brain areas associated with executive control and norm-compliance (i.e., dACC, LPFC).

To test these hypotheses, we developed a paradigm that combines an interpersonal transgression task that induces guilt (Koban et al., 2013; Yu et al., 2014) with a minimal group manipulation that induces group identity (Dunham, 2018; Otten, 2016). In this paradigm, we manipulated participants' relationship with transgressors who were from the in-group/out-group with participants and whether the participants themselves committed the transgression, or they just observed the in-group/out-group transgressors commit the transgression. Specifically, participants either observed two in-group (*In-group_ Observe*) or two out-group (*Out-group_ Observe*) members cause harm to an anonymous victim group, or participants themselves together with either an in-group (*In-group_ Commit*) or an out-group (*Out-group_ Commit*) member directly caused harm to the victims. Then they rated their level of guilt (Experiment 1) or divided 20 *yuan* (~3 USD) between themselves and the victim group (Experiment 2). The monetary allocation decision, which the participants believed was unknown to the victim group, was included as a measure of reparative motivation and has been shown to be a reliable indicator of guilty feeling (Ketelaar and Au, 2003; Gao et al., 2018; Yu et al., 2014). We predicted that group-based guilt induced by observing in-group (relative to outgroup) members cause harm to another, would manifest in self-reported guilt, compensatory behavior, and brain activations in areas associated with personal guilt. In addition, we adopted conjunction analysis and multivariate pattern analysis (MPVA) to formally examine whether personal and group-based guilt exhibit shared or distinct neural representations (Experiment 2). If group-based guilt is built on the core cognitive-affective processes of personal guilt, we should not only observe overlapping univariate activations by these two types of guilt, but also predictive power of a personal guilt MVPA classifier generalizable to predict group-based guilt.

## 2. Materials and methods

### 2.1. Participants

For Experiment 1 (behavioral experiment), we recruited 24 participants (12 females, mean age 19.3 ± 0.8 years). For Experiment 2 (fMRI experiment), thirty-five right-handed participants completed the experiment, four of which were excluded due to excessive head motion (>3 mm), leaving 31 participants (19 females, mean age 21.3 ± 1.1 years) in data analysis. Three participants recruited for Experiment 2 did not complete the experiment due to expressed doubts about the experimental setup. None of the participants reported any history of neurological or psychological disorders. Written informed consent was obtained from every participant before the experiments. This study was carried out in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychological and Cognitive Science, Peking University.

### 2.2. Experimental design and procedures

#### 2.2.1. Procedures of Experiment 1 (behavioral experiment)

*2.2.1.1. Overview.* In the current experiment, six participants of the same sex were recruited on each experimental session (none of them had known one another before the session). Upon arrival, participants were told that all six of them were predetermined to be assigned to group A (Transgressor-group) and six other co-players (confederates of the experimenter) in another room were assigned to group B (Victim-group). The task consisted of two phases. In the first, minimal group manipulation phase, the six participants of group A were randomly divided into two sub-groups of three members each to build in-group/out-group context; in the second phase, the participants played multiple rounds of a dot-estimation game either with two in-group partners or two out-group partners. The victims would receive electric shocks depending on the performance of the participant and/or other Role A players (Fig. 1). The participants were explicitly told that the victims could not reciprocate the electric shocks, and that the participants' identity would not be revealed to any other players (i.e., in-group and out-group players and the victim group) during or after the experiment, neither would they meet in person. This was to ensure anonymity and minimize reputation concerns.

*2.2.1.2. Minimal group manipulation.* In the first phase, the six participants of group A were randomly divided into two sub-groups of three members each (a "Yellow Group" and a "Blue Group"). They were asked to wear a yellow or a blue T-shirt corresponding to their group membership. Each sub-group was required to work together to solve a "winter survival problem" (Johnson and Johnson, 1991) to enhance group identity. The background of the winter survival problem is that the participants took plane that had crash-landed in the woods of a

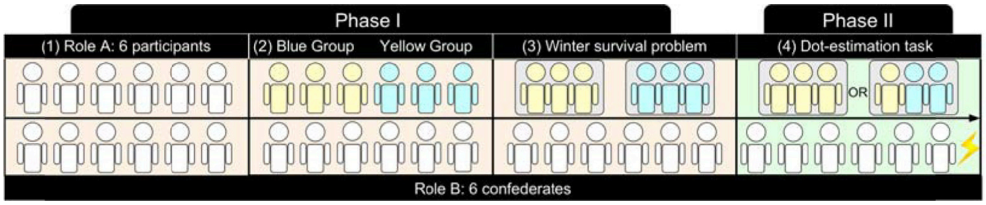


**Fig. 1. Overview of the intergroup game.** Phase I was for the groups formation and Phase II was the dot-estimation task for guilt induction. Phase I had three stages. (1) *Role assignment*. Six same-sex students were recruited each time and were assigned Role A. They were told that another 6 participants (confederates, the second line of stick-figure in the figure) in another room were assigned Role B. (2) *Group assignment*. The six participants in Role A were further divided into a 'Yellow Group' and a 'Blue Group' to induce the in-group/out-group context. They were told that the participants in Role B were also divided into two 3-member groups. (3) *Group membership formation and reinforcement*. To strengthen the identification with the newly formed minimal group, each group was asked to work together to solve a 'winter survival problem' within 6 min. Phase II consisted of one stage, namely the *Dot-estimation task*. Here, the participants (indicated by the left-most yellow figure in the three stick-figures) performed a dot-estimation task with two in-group partners (e.g., the two yellow members in the example) or two out-group partners (e.g., the two blue members in the example). A group from Role B (i.e., the victims) would receive a painful shock if Role A (i.e., the transgressors) failed. Color should be used for Fig. 1 in print.

northeastern region in mid-January, and they were required to rank-order 10 items salvaged from the broken plane (a lighter, a chocolate bar, a gun, newspaper, etc.) according to their importance for survival. The 3 individuals needed to discuss the problem together and reach a single consensus ranking within 6 min. As a manipulation check, we asked the participants to complete a scale of psychological distance to their group (a modified version of Inclusion of Others; Aron et al., 1992) and a 6-item questionnaire of group identity (example items: "How much do you identify with the Yellow group?" and "To what extent do you feel strong ties with the Yellow group?"; Falk et al., 2014) immediately after the mini-group manipulation. Player As were explicitly told that player Bs were also divided into two 3-member groups.

*2.2.1.3. Pain calibration.* After the group manipulation, participants were told that the victims (i.e., the confederates) would receive pain stimulation if the participants failed in their dot-estimation task (see below). To familiarize participants with the pain stimulation, all the participants underwent a pain calibration procedure. An intra-epidermal needle electrode was attached to the left wrist of participants for cutaneous electrical stimulation (Inui et al., 2002). Pain calibration begun with 8 repeated pulses (with each pulse having 0.2 mA and lasted for 5 ms with a 10 ms inter-pulse interval). Then we gradually increased the intensity of each pulse until participants reported 7 on a scale from 1 ('not painful') to 8 ('intolerable'). Note that an 'intolerable' shock here meant that the participants did not want to experience that shock anymore. We made this clear to the participants before the calibration procedure. All participants reported that the pain stimulation rated as 7 was really painful. They were told that the victims underwent the same pain calibration procedure and would receive the pain stimulation they rated as 7 if the transgressors failed in the dot-estimation task. Notably, no one would in fact receive pain stimulation due to participants' performance, and the so-called victims were in fact confederates of the experimenters. To further protect the participants and ensure ethical practice, each time before we increased the intensity of pain, we would always ask the participants' consent. We would only increase the pain intensity or deliver any pain stimulation with participants' consent, and would stop the procedure whenever the participants decided so.

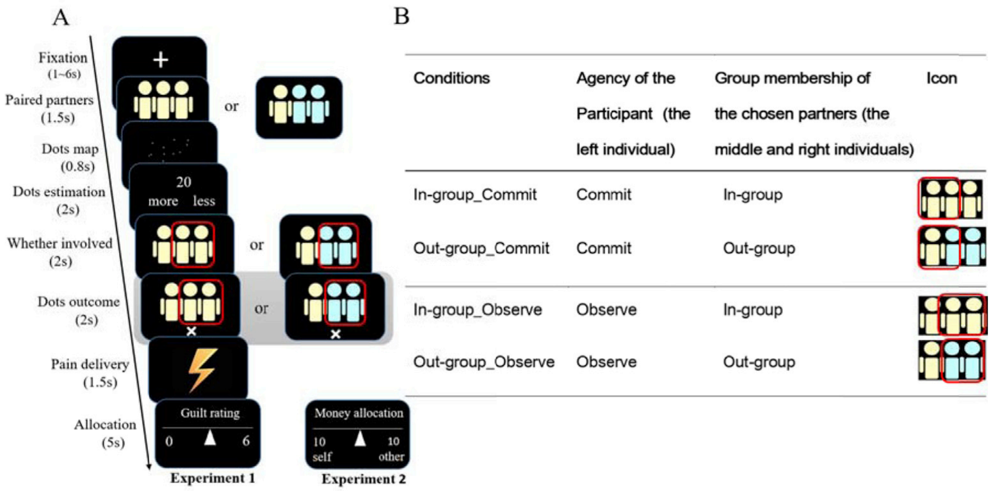
*2.2.1.4. Dot-estimation task.* In this task, each round began by informing the participant (represented by the left of the three stick-figures in the screens with the figures in Fig. 2A) whether the two partners he/she paired with in the current round (represented by the middle and right of the three stick-figures in the screens of Fig. 2A) were from the "Yellow Group" or the "Blue Group". Each of the three players was required to estimate the number of dots presented on the screen, press a corresponding button to indicate whether their estimate was 'More' or 'Less' than a reference number presented on the screen (randomly chosen from 19, 20, 21, and 22), and then press a button to confirm their choice. If the participant failed to confirm his/her choice within 2 s, the current trial would start again with a different dots map presented. The participants were explicitly told that the average accuracy for the dot-estimation task is 75%, to make them believe that they could estimate correctly. Two out of the three responses in the current round were randomly selected, as indicated by one/two red rectangles, and the outcome (success or failure) was presented on the next screen (see below for details on the different combinations of response selections). If the chosen estimates were both correct (i.e., filler trials), an 'O' sign would appear on the screen indicating that the current trial was successful and no painful stimulation would be delivered, and the current trial terminated there. If one or both of the selected estimations were incorrect (i.e., experimental trials, Fig. 2B), a ' × ' sign indicating failure of the current round was presented and one victim group of group B would be randomly selected to receive pain stimulation, as indicated by a shock sign appearing on the screen. Then, the participant was asked to rate his/her level of guilt on a 0–6 scale (in increments of 1) by pressing a key to increase or decrease the rating before pressing the space bar to confirm his/her choice. The software Presentation was used to display the stimuli and collect the data.

The experimental trials (i.e., failure trials) consisted of the four combinations of the two experimental factors, namely, Group membership (In-group *vs.* Out-group) and Agency (Commit *vs.* Observe), forming a 2 × 2 within-participant design. 'Group membership' refers to the group membership of the two partners and 'Agency' refers to whether the participant's performance was selected and therefore directly involved in causing harm to the victims. Therefore we had four experimental conditions (Fig. 2B): 1) *In-group_ Observe*, where the performance of the two in-group members were selected, and the performance of the participant was not selected; 2) *Out-group_ Observe*, where the performance of the two out-group members were selected, and the estimation of the participant was not selected; 3) *In-group_ Commit*, where the performance of one in-group member and that of the participant were selected; and finally 4) *Out-group_ Commit*, where the performance of one out-group member and that of the participant were selected. Similarly, there were 4 possible combinations for success trials corresponding to the four experimental conditions and they were the filler trials. Unbeknownst to the participants, the outcome was predetermined by a computer program, ensuring that all the conditions had the same number of trials.

The experiment consisted of 84 trials (16 for each experimental condition, and 5 for each filler combination). Trials were presented in a pseudorandom order to each participant with the constraint that no more than three consecutive trials were from the same condition. After the experiment, each participant rated on a 9-point Likert scale (1 = 'not at all', 9 = 'very strong'), indicating his/her responsibility, fear, and anger in the four experimental conditions, and received 50 yuan (~7.7 USD) for their participation.

*2.2.2. Procedures of Experiment 2 (fMRI)*

Procedures of Experiment 2 were identical to those of Experiment 1, except that 1) one participant was recruited each time for scanning, and the participant met five confederates (2 male, 3 female, 23.6 ± 1.3 years)



| Conditions | Agency of the Participant (the left individual) | Group membership of the chosen partners (the middle and right individuals) | Icon |
|---|---|---|---|
| In-group_Commit | Commit | In-group | |
| Out-group_Commit | Commit | Out-group | |
| In-group_Observe | Observe | In-group | |
| Out-group_Observe | Observe | Out-group | |

**Fig. 2. Experimental design and procedure. (A)** Each trial began by informing the participants (indicated by the left most yellow stick-figure) whether they were paired with in-group or out-group partners on the current round (*Paired partners*). Then a number of dots distributed randomly on the screen were presented (*Dots map*). The participants needed to quickly estimate the number of dots and compare the estimation with a specific number (e.g., 20) to be appeared on the next screen (*Dots estimation*). Then, the performance of two out of the three players were randomly chosen, as indicated by a red rectangle (*Whether involved*). On the next screen, the outcome (i.e., correct vs. incorrect) would be presented (*Dots outcome*). If one or two of the chosen estimations were incorrect, a ' × ' signal (indicating the failure of the current round) would be presented, and one group from Role B was randomly chosen to receive pain stimulation. This, if happened, was indicated by a lightening sign (*Pain delivery*). After pain delivery, the participant rated their feelings of guilt on a 7-point Likert scale (Experiment 1) or allocated a portion of 20 *yuan* to victims (Experiment 2). The critical event for fMRI data analysis was the *Dots outcome* screen (gray rectangle), where the participants received the information about the harm of the victim, their own causal contribution to the harm, and the group membership of the other player(s) chosen. **(B)** Experimental conditions and their corresponding icons presented to the participants. Note that our conditions of interest are those in which the players failed and caused pain to the victim group. *In-group_ Observe*, where the performance of the two in-group members was selected, and the performance of the participant was not selected; *Out-group_ Observe*, where the performance of the two out-group members was selected, and the estimation of the participant was not selected; *In-group_ Commit*, where the performance of one in-group member and that of the participant were selected; *Out-group_ Commit*, where the performance of one out-group member and that of the participant were selected.

upon arrival at the laboratory; 2) at the end of each estimation-failure round, instead of rating feelings of guilt, the participant was asked to divide 20 *yuan* (~3 USD) between him/herself and the 3 player Bs who received the pain stimulation, with the knowledge that the player Bs were unaware of the existence of this money distribution. The monetary allocation decision was included as a measure of reparative motivation and has been shown to be a reliable readout of guilty feeling (Ketelaar and Au, 2003; Yu et al., 2014). Thus, the amount allocated to the player Bs was interpreted as a measure of compensation for electric shocks. After scanning, participants rated their responsibility, fear, anger, and guilt on a 9-point Likert scale (1 = not at all, 9 = very strong) for each of the four experimental conditions. At end of Experiment 2, the participants were explicitly told that their allocation on one trial would be randomly selected and made real as an extra payment at the end of the experiment. Thus, the participants' final payoff was the baseline payoff (100 *yuan,* approximately 15 USD) and the additional amount of allocation the participants kept for themselves.

### 2.2.3. Direct replication of the behavioral findings of experiment 2

To confirm the stability of the behavioral patterns observed in the fMRI experiment, we performed a behavioral experiment (Experiment 3) with the same procedures as the fMRI experiment in an independent sample of 36 participants, but with 6 same-sex participants recruited each time as in Experiment 1. One participant was excluded due to a technical error, leaving 35 (23 female, mean age 21.7 ± 0.9 years) for data analysis.

### 2.2.4. Debriefing the participants

In the current study, we explicitly told the participants that the average accuracy for the dot-estimation task is 75% in order to make them more engaged in the task. The dots stayed on the screen for a very short period of time (~0.8 s), therefore it is almost impossible for the participants to be entirely sure whether their performance was correct. Although it is possible that at the end of the experiment the participants may have noticed that their overall accuracy was a bit lower than 75%, no one expressed doubt that the feedback might have been predetermined. When it comes to participants' belief about the experimental setup, the post-experiment interview showed that no one expressed doubts about the existence of the victims and other players. Specifically, in the post-experiment interview, we asked participants whether they believed the existence of the victims (e.g., "Do you think the victims would receive pain stimulation if you or other players failed in the task?") and whether they believed that they were performing the task with other players (e.g., "Do you think you performed the task by yourself or with other group members?"). Among the participants who completed the task, all gave confirmative responses to these questions. After the post-experiment interview, we debriefed the participants and they expressed relief after knowing that no one was actually hurt.

## 2.3. Statistical analysis

### 2.3.1. Behavioral data analysis

The behavioral data were analyzed using linear mixed effects models implemented in the R software environment with the lmer4 package (Baayen et al., 2008; Bates et al., 2014). To check how the Group (In-group *vs.* Out-group) and the Agency (Commit *vs.* Observe) modulated the guilt rating (Experiment 1) or compensation behavior (Experiments 2 and 3), the current LMM included two fixed-effect variables (Group and Agency) and their possible interactions, and one random factor (participants). To control for any potential confounding effects, we added all fixed factors (Group, Agency, and Group × Agency) into random slopes to better generalize the LMM analysis (Barr et al., 2013).

For all the mediation analyses in our study, we bootstrapped the mediating effect 20,000 times using the SPSS version of the INDIRECT macro (http://www.afhayes.com/) developed by Preacher and Hayes (2008) and obtained the bias-corrected 95% confidence interval of the indirect effects.

### 2.3.2. Imaging data acquisition

Imaging data were acquired using a Siemens 3.0 T Prisma scanner at the Beijing MRI Centre for Brain Research at Peking University (China). T2*-weighted echo-planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast were collected in 33 axial slices parallel to the anterior-posterior commissure line to ensure coverage of the whole cerebrum (matrix 64 × 64, in planar resolution). Images were acquired in an interleaved order with no inter-slice gap (TR = 2000 ms, TE = 30 ms, voxel size = 3.5 mm × 3.5 mm × 3.5 mm, field of view = 224 mm × 224 mm, flip angle = 90°). A high-resolution, whole-brain structural scan (1 × 1 × 1 mm$^3$ isotropic voxel MPRAGE) was acquired before functional imaging.

### 2.3.3. Imaging data preprocessing

The fMRI images were preprocessed using Statistical Parametric Mapping software SPM8 (Wellcome Trust Center for Neuroimaging, London, UK). The first five volumes of each run were discarded to allow for stabilization of magnetization. The remaining images were slice-time corrected, motion-corrected, re-sampled to 3 × 3 × 3 mm$^3$ isotropic voxels, normalized to the Montreal Neurological Institute (MNI) space, spatially smoothed using an 8-mm full width at half maximum Gaussian filter, and temporally filtered using a high-pass filter with 1/128 Hz cutoff frequency.

### 2.3.4. Whole-brain general linear model analyses

Whole-brain exploratory analysis based on the general linear model was conducted first at the participant level, and then at the group level. To examine the neural responses to transgression outcomes, the data analysis focused on the brain responses associated with the presentation of the dot estimation outcomes. At the participant-level statistical analysis, failed dots-estimation outcomes corresponding to the four experimental conditions and 18 other regressors were separately modeled in a General Linear Model (GLM). Separate regressors in GLM were specified for fMRI responses to:

- R1: Combined regressor of no interest (duration = 4.3 s), which consisted of the presentation of the in-group/outgroup partners whom the participant paired with ("paired partners" screen), the dot map, and participants' dot-estimation responses ("dot estimations" screens);
- R2-R5: Cue for involvement screen (duration = 2 s), which was indicated by a red rectangle around 2 of the 3 silhouettes. Each of the four experimental conditions was modeled in a separate regressor;
- R6-R9: Failed dot-estimation outcomes (2 s), separately modeled for each of the four experimental conditions;
- R10-R13: Successful dot-estimation outcomes (2 s), separately modeled for each of the four filler conditions;
- R14: The presentation of pain shock cue (1.5 s);
- R15: Allocation screen, which required the participant to allocate 20 yuan between him/herself and the victims (5 s);
- R16: Missed trials (4.3 s);
- R17-R22: Six head motion parameters, which were modeled separately in the GLM to account for artifacts in the scanner signal.

All regressors were convolved with a canonical hemodynamic response function. At the group-level, the four beta maps corresponding to the four experimental conditions of each transgressor were fed into a flexible factorial design matrix.

At the group-level, we defined the following contrasts:

- group-based guilt: *In-group_ Observe* > *Out-group_ Observe*;
- personal guilt: *Out-group_ Commit* > *Out-group_ Observe*;
- main effect of Agency: (*In-group_ Commit* + *Out-group_ Commit*) > (*In-group_ Observe* + *Out-group_ Observe*);
- main effect of group membership: (*In-group_ Commit* + *In-group_ Observe*) > (*Out-group_ Commit* + *Out-group_ Observe*);

We focused on the group-based guilt contrast (*In-group_ Observe* > *Out-group_ Observe*) and the personal guilt contrast (*Out-group_ Commit* > *Out-group_ Observe*) as they were the main concern of our study. We took the simple effect contrast '*In-group_ Observe* > *Out-group_ Observe*' as the defining contrast for group-based guilt because the victim's harm as well as participants' causal contribution to the harm were identical in these two conditions; the only difference was participants' relationship with the transgressor. In the current study, 'personal guilt' was defined by the contrast '*Out-group_ Commit* > *Out-group_ Observe*'. It should be noted that the contrast '*In-group_Commit* > *In-group_Observe*' also minimizes the impact of group membership. In that sense, we could define personal guilt just as well with this contrast. However, we also aimed to maximize participants' perceived causal contribution or responsibility to the transgression in the definition of personal guilt. Based on the behavioral data, the difference in responsibility was larger in the contrast 'Out-group_Commit > Out-group_Observe' than in the contrast 'In-group_ Commit > In-group _ Observe'. We therefore defined 'personal guilt' with the contrast '*Out-group_ Commit* > *Out-group_ Observe*'.

The statistical threshold for the whole-brain exploratory analysis was defined as $P < 0.005$ uncorrected at peak level with cluster size ≥46 voxels. The reason for us to choose this threshold was threefold: First, as the first neuroimaging study on the neural basis of group-based guilt, the current study, especially the whole-brain univariate analysis, has an exploratory aspect. It is therefore important to balance Type I/Type II error (Lieberman and Cunningham, 2009). A stringent threshold may hinder interesting and meaningful findings from emerging. These initial findings could then be replicated and extended by future studies inspired by the initial findings. Neural results related to social-affective processes are particularly vulnerable to Type 2 error given their relatively larger inter-individual variability compared with, for example, perceptual processes. Therefore, based on the recommendations in Lieberman and Cunningham (2009), we set the intensity (i.e., voxel-level) threshold for the whole-brain univariate analysis to be $p < 0.005$. Second, to buttress the univariate exploratory analysis, we carried out a hypothesis-driven, independent ROI analyses, which lent support to our whole-brain analysis and circumvent the potential impact of Type 1 error. And third, our conclusion that group-based guilt and personal guilt share a core neurocognitive process is further supported by a multivariate analysis, which did not rely on the magnitude of activation of individual voxels. Given the three pieces of independent and converging evidence, we believed the risk of obtaining false alarms in the whole-brain univariate analysis was minimized.

### 2.3.5. Conjunction analyses

To identify brain areas that are shared by personal guilt and group-based guilt, we performed a conjunction analysis (Price and Friston, 1997) over the personal guilt contrast (*Out-group_ Commit* > *Out-group_ Observe*) and the group-based guilt contrast (*In-group_ Observe* > *Out--group_ Observe*). Conjunction analysis allowed us to combine these two comparisons to look for areas shared by personal guilt and group-based guilt while simultaneously eliminating areas activated in only one of the comparisons. Thus, this approach both increases statistical power (relative to only looking at, for example, the personal guilt contrast or the group-based guilt contrast), while also eliminating comparison specific activations which may reflect idiosyncratic influences of one of these two contrasts (Price and Friston, 1997). This conjunction was formulated as conjunction null hypothesis (Friston et al., 2005; Nichols et al., 2005) and should therefore only yield activations that are significantly present in both original contrasts of the conjunction. The null hypothesis for "conjunction null hypothesis" is that "not all contrasts activated this voxel." If the conjunction results are significant, the null hypothesis is rejected and the conclusion is that "all contrasts activated this voxel." That is, conjunctions represent a logical 'and', requiring both contrasts to be separately significant for the conjunction to be significant.

### 2.3.6. Multivariate pattern analysis of imaging data

Compared to the univariate analysis, the Multivariate pattern analysis (MVPA) could increase the amount of information that can be decoded from brain activity (i.e. spatial pattern). Thus, a supplementary MVPA analysis was carried out in the conjunction region to check whether the spatial pattern of group-based guilt was similar to those of personal guilt.

Our rationale for training the classifier on personal guilt is that we take personal guilt as a prototypical species of guilt that exemplifies the core cognitive-affective processes present in a family of emotions that fall into the category of guilt (e.g., survivor guilt, group-based guilt, guilt for failure in personal goals and so on) (Deigh, 1999; Schoeman, 1987; Shaver et al., 1987). The reason that those tokens of emotion are labeled as "guilt", both by emotion researchers and in everyday discourse, is because they share those core cognitive-affective processes exemplified by personal guilt. In that sense, we treat group-based guilt as a variant on the theme of personal guilt, therefore logically it makes more sense to train the classifier based on a more encompassing form of guilt and apply it to a more specific type of guilt. We used linear Support Vector Machine (SVM) (Friedman et al., 2001; Wager et al., 2013) to train a multivariate pattern classifier on personal guilt trials and apply the classifier to discriminate group-based guilt. With a leave-one-out cross-validation method, we calculated the accuracy of the SVM classifiers using the forced-choice test (Chang et al., 2015; Wager et al., 2013; Woo et al., 2014). We also calculated the accuracy for *In-group_ Observe vs. Out-group_ Observe*. To be noted that, in case the SVM effects were driven solely by the effects of response amplitude already observed in the GLM analysis, the mean univariate response magnitude in the overlapped region was subtracted (Coutanche, 2013; Smith et al., 2011).

## 3. Results

### 3.1. Group-based guilt elicited by an interaction-based minimal group paradigm

As a manipulation check, we first examined whether the participants felt closer to the in-group than to the out-group members. In Experiment 1, paired sample *t*-test showed that the participants indeed felt closer to and had stronger identity with the in-group partners than the out-group partners (Table 1): closeness, $t\ (23) = 10.0$, $p < 0.001$, $d = 2.08$; identity, $t\ (23) = 9.8$, $p < 0.001$, $d = 1.99$. This was replicated in Experiment 2: The participants felt closer to and identified more with the in-group partners than the out-group partners (Table 1): closeness, $t\ (30) = 7.8$, $p < 0.001$, $d = 1.41$; identity, $t\ (30) = 6.4$, $p < 0.001$, $d = 0.77$. These results demonstrated the validity of in-group/out-group manipulation.

Did participants feel guiltier and allocate more money to the victims as compensation when they are causally involved in the victims' harm? More importantly, did group membership of the agents who caused the victims' harm play a role in participants' guilt and compensation when they merely observed but not caused the harm? To answer these questions, we examined the patterns of self-reported guilt (Experiment 1) and monetary allocation (Experiment 2) using linear mixed effects. Not surprisingly, we found that in Experiment 1 participants felt guiltier when they committed the harm than when they merely observed, $\beta = 0.68$, $SE = 0.08$, $t = 8.50$, $p < 0.001$. More interestingly and consistent with our hypothesis, participants expressed more guilt when they observed in-group partners cause the harm than when they observed out-group partners cause the same harm, $\beta = 0.29$, $SE = 0.06$, $t = 4.90$, $p < 0.001$. Such difference was reduced when comparing the two *Commit* conditions, supported by a significant Group membership × Agency interaction, $\beta = 0.08$, $SE = 0.04$, $t = 2.26$, $p = 0.03$ (Table 2; Fig. 3A). Monetary allocation in Experiment 2 showed a similar pattern as the self-reported guilt ratings in Experiment 1 (Fig. 3B). In general, participants compensated more when they themselves committed the harm than when they merely observed the harm, $\beta = 0.73$, $SE = 0.09$, $t = 7.44$, $p < 0.001$. More specifically, the participants allocated more when they observed in-group partners cause the harm than when they observed out-group partners cause the same harm, $\beta = 0.38$, $SE = 0.11$, $t = 3.53$, $p < 0.001$. Similar to the pattern of self-reported guilt, such difference was reduced when comparing the two *Commit* conditions, supported by a significant Group membership × Agency interaction, $\beta = 0.16$, $SE = 0.08$, $t = 2.14$, $p = 0.04$ (Table 2; Fig. 3B). Other contrasts and statistic details of these regression analysis can be found in the *Supplementary Results of Experiments 1 and 2*.

To examine the possible effects of participants' sex on group-based guilt, we also carried out a Group membership (In-group vs Out-group) by Agency (Commit vs Observe) by Sex (male vs female) mixed effect ANOVA for guilt rating (Exp. 1) and monetary allocation (Exps. 2&3), to check whether guilt rating and monetary allocation were modulated by participants' sex. The three-way interaction was not significant in any of our 3 experiments (for guilt ratings: Exp. 1: F (1, 22) = 1.04, p = 0.32; for allocation: F (1, 29) = 0.15, Exp. 2: p = 0.70; Exp. 3: F (1, 33) = 0.29, p = 0.58). Participants' sex did not show significant main effects either.

### 3.2. Shared responsibility explains group-based guilt and compensation

To investigate the cognitive processes underlying group-based guilt, we examined the role of shared responsibility in group-based guilt. Not surprisingly, participants perceived higher responsibility in the *Commit* conditions than in the *Observe* conditions ($F\ (1, 23) = 151.17$, $p < 0.001$, $\eta^2_p = 0.87$ for Experiment 1; $F\ (1, 30) = 79.30$, $p < 0.001$, $\eta^2_p = 0.73$ for Experiment 2). Importantly, this effect was modulated by partners' group membership: the interaction between group membership (In-group *vs.* Out-group) and Agency (Commit *vs.* Observe) was significant for both experiments ($F\ (1, 23) = 7.55$, $p = 0.011$, $\eta^2_p = 0.25$ for Experiment 1; $F\ (1, 30) = 5.45$, $p = 0.03$, $\eta^2_p = 0.15$, Experiment 2; see Table 2 for details). Specifically, pairwise comparisons showed that the participants felt more responsible in the *In-group_ Observe* condition than in the *Out-group_ Observe* condition ($F\ (1, 23) = 11.38$, $p = .003$, $\eta^2_p = 0.33$ for Experiment 1 and $F\ (1, 30) = 13.87$, $p < 0.001$, $\eta^2_p = 0.32$ for Experiment 2). The fact that they had no objective responsibility in either case suggests that the participants 'inherited' moral responsibility for transgressions committed by in-group members. Moreover, the difference in perceived responsibility between observing in-group versus out-group transgression was positively correlated both with the difference in self-reported guilt rating (Experiment 1, $r = 0.45$, $p = 0.03$; Experiment 2, $r = 0.52$, $p = 0.003$) between these two conditions. This indicated that the perceived moral responsibility was associated with the experience of group-based guilt. No significant effect was found for fear and angry emotion. The relationship between perceived moral responsibility in self-reported guilt was replicated in Experiment 3 (see *Supplementary Results of Experiments 3*).

Research on personal guilt has suggested that guilt functions as an intermediate state between acknowledging responsibility of transgression and reparative behavior (e.g., Yu et al., 2014). Here we provided more concrete evidence for this conjecture by a mediation analyses (Preacher and Hayes, 2008). We found a significant indirect path from perceived responsibility *via* self-reported guilt to monetary allocation (mediating effect estimate = 0.19, SE = 0.09, 95% confidence interval was [0.009, 0.357], see *Supplementary Results of Experiments 1 and 2* for details). Importantly, for group-based guilt, we found a similar indirect pathway *via* guilt (the mediating effect estimate = 0.18, SE = 0.08, 95% confidence interval [0.003, 0.331] (Fig. 3D), see *Supplementary Results of Experiments 1 and 2* for details). This finding lends support to the Intergroup Emotion Theory, according to which group-based guilt should function in a similar way as personal guilt in mediating the relationship between perceived responsibility and compensation behaviors.

**Table 1**
Results of manipulation check.

| Item | In-group | Out-group | *t*-value |
|---|---|---|---|
| Experiment 1 | | | *t* (23) |
| Closeness | 4.7(.2) | 2.7(.2) | 10.0 *** |
| Group identity | 5.4(.2) | 3.5(.2) | 9.8 *** |
| Experiment 2 | | | *t* (30) |
| Closeness | 4.4(.3) | 2.7(.2) | 7.8 *** |
| Group identity | 4.3(.2) | 3.4(.2) | 6.4 *** |

*Note.* Standard errors (*SEs*) are shown in parentheses. Significant paired sample *t*-test is denoted by * $p < .05$, ** $p < .01$, *** $p < .001$.

**Table 2**
Behavioral results in Experiments 1 and 2.

| Item | In-group_Commit | Out-group_Commit | In-group_Observe | Out-group_Observe | Interaction *T/F* |
|---|---|---|---|---|---|
| Experiment **1** | | | | | |
| Online measure | | | | | Interaction *T* |
| Guilt rating | 4.0 (.1) | 3.6 (.1) | 2.8 (.1) | 2.1 (.1) | 2.26* |
| Post-experiment measures | | | | | Interaction *F*(1, 23) |
| Responsibility | 6.8 (.3) | 6.6 (.4) | 4.5 (.5) | 3.3 (.4) | 7.55* |
| Fear | 3.5 (.5) | 3.2 (.6) | 3.1 (.5) | 2.2 (.3) | 2.47 |
| Angry | 3.5 (.4) | 2.9 (.5) | 2.8 (.4) | 2.4 (.4) | 0.10 |
| Experiment **2** | | | | | |
| Online measure | | | | | Interaction *T* |
| Monetary allocation | 13.5(.2) | 13(.2) | 12.3(.2) | 11.2(.2) | 2.14* |
| Post-experiment measures | | | | | Interaction *F*(1, 30) |
| Responsibility | 6.9 (.3) | 6.7 (.3) | 4.4 (.4) | 3.3 (.4) | 5.41* |
| Guilt | 6.5(.3) | 5.9 (.4) | 4.1 (.4) | 3.2 (.4) | 1.05 |
| Fear | 3.6 (.3) | 2.7 (.4) | 3.3 (.3) | 2.6 (.3) | 0.16 |
| Angry | 3.1 (.4) | 2.7 (.4) | 3.0 (.3) | 2.6 (.3) | 0.11 |

*Note.* Standard errors (*SEs*) are shown in parentheses. *SEs* of Online measures of Experiment 1 and Experiment 2 were estimated by data from every single trial under each condition. *SEs* of Post-experiment measures were standard errors of means. Significant two-way interaction is denoted by * $p < .05$.

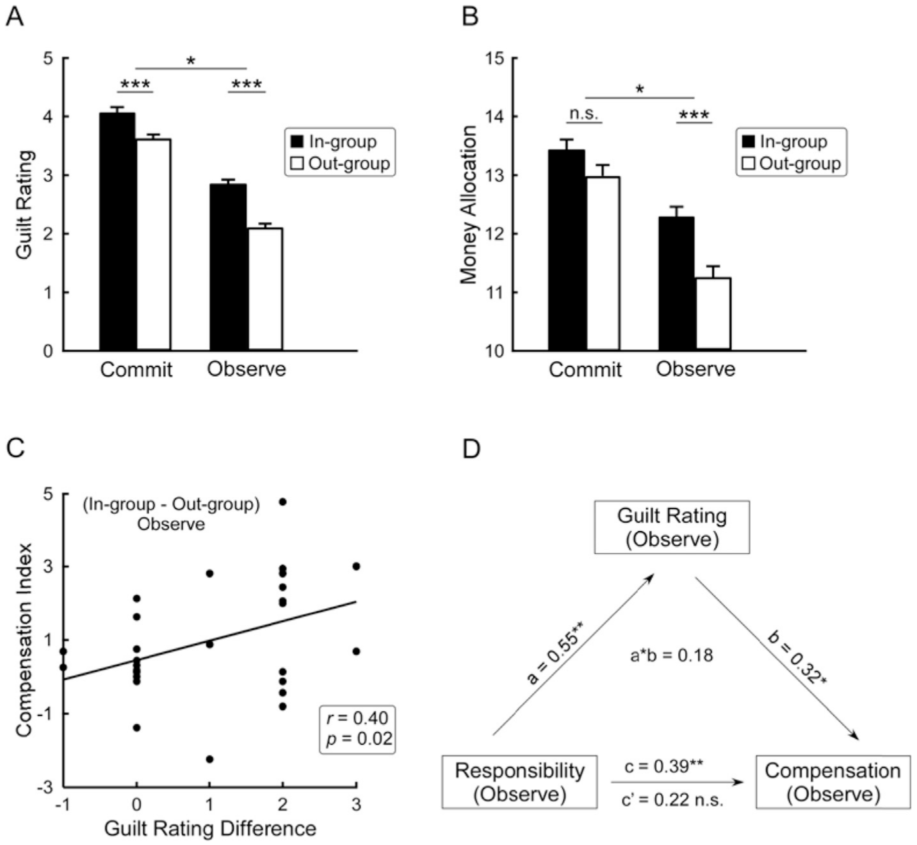


**Fig. 3. Behavioral results of Experiment 1 (A) and Experiment 2 (B).** For Fig. 2A and B, asterisks on the top indicate significant Group (In-group *vs.* Out-group) by Agency (Commit *vs.* Observe) interaction. Asterisks below indicate significance in post hoc test. **(C)** In Experiment 2, the post-experiment guilt rating difference (Observe: *In-group_ Observe* > *Out-group_ Observe*) positively correlated with the corresponding allocation difference. **(D)** The indirect pathway from the shared responsibility, via guilt rating, to compensation in Experiment 2. Error bars are standard errors estimated for data in each condition. ***$p < .001$, **$p < .01$, *$p < .05$.

### 3.3. *Brain activations associated with personal and group-based guilt*

Here we presented the brain activation patterns revealed by the contrasts hypothesized to reflect group-based and personal guilt, respectively. The activation patterns corresponding to the main effects of Agency and Group can be found in *Supplementary Neuroimaging Results.*

To identify brain regions associated with group-based guilt, we focused on brain responses associated with the outcome feedback of dot estimation. We defined the critical contrast "*In-group_ Observe* > *Out-group_ Observe*", which corresponds to the effect of group-based guilt. This contrast revealed activations in anterior middle cingulate cortex (aMCC; MNI coordinates = [6, 26, 28]; $k$ = 85 voxels) and right anterior insula (AI; MNI coordinates = [27, 20, −11]; $k$ = 78 voxels) (Fig. 4A). aMCC and right AI have been consistently implicated in imagining and experiencing personal guilt (Chang et al., 2011; Yu et al., 2014) and negative self-evaluation in social contexts (Immordino-Yang et al., 2009; Kédia et al., 2008; Koban et al., 2013; Sanfey et al., 2003; Zaki et al., 2007). To illustrate the activation patterns, we extracted the regional parameter estimates from 27 voxels around the peak coordinates at aMCC and right AI. The parameter estimates extracted from aMCC (Fig. 4B) and right AI (Fig. 4C) exhibited a pattern similar to the pattern of monetary allocation. Moreover, aMCC activation difference (*In-group_ Observe* > *Out-group_ Observe*) positively correlated with the post-scan guilt rating difference between these two conditions ($r$ = 0.45, $p$ = 0.011), indicating that the aMCC was involved in the processing of group-based guilt.

We next examined whether group-based guilt shared a similar neurocognitive process with personal guilt. In the current study, ‘personal guilt’ was defined by the contrast ‘*Out-group_ Commit* > *Out-group_ Observe*’. This contrast, while keeping the impact of group membership to its minimal, captured the difference in the participants’ causal contribution to the transgression, thereby reflecting neural processing of personal guilt. Replicating previous neuroimaging findings about personal guilt (Koban et al., 2013; Yu et al., 2014), this contrast (*Out-group_ Commit* > *Out-group_ Observe*) revealed the activations in aMCC (MNI coordinates = [12, 17, 40]; $k = 153$ voxels) and supplementary motor area (SMA) (MNI coordinates = [15, 5, 67]; $k = 62$ voxels) (Fig. 5A).

We extracted the regional parameter estimates from 27 voxels around the peak coordinates of aMCC (Fig. 5B) and further demonstrated that the aMCC activation difference (*Out-group_ Commit* > *Out-group_ Observe*) was positively correlated with the difference in self-reported guilt between these two conditions ($r = 0.43$, $p = 0.02$). The right AI has been implicated in representing personal guilt (Kédia et al., 2008; Koban et al., 2013; Yu et al., 2014). We extracted the regional parameter estimates from 27 voxels around the peak coordinates of an independently defined right AI region of interest ([36, 30, −8], coordinates reported in Yu et al., 2014). The activations within this ROI (Fig. 5C) showed a similar pattern with that of the aMCC. Statistical analysis further confirmed that for this ROI, the activation in the *Out-group_ Commit* condition is stronger than in the *Out-group_ Observe* condition, $t = 2.75$, $p = 0.01$.

### *3.4. Group-based guilt shares brain representation with personal guilt*

The aMCC was implicated in the whole-brain contrasts both for the group-based guilt contrast (*In-group_ Observe* > *Out-group_ Observe*) and for the personal guilt contrast (*Out-group_ Commit* > *Out-group_ Observe*). This suggested the possibility of a shared neuropsychological processes underlying these two forms of guilt. To examine this in a more principled way, we first performed a conjunction analysis between the aforementioned contrasts. The overlapping region identified from the conjunction analysis confirmed that the overlapped aMCC ((MNI coordinates = [6, 26, 28]; $k = 31$ voxels), Fig. 6A) was sensitive to both personal guilt and group-based guilt. Moreover, we performed a multivariate pattern analysis (MVPA) as a supplementary analysis in the overlapped aMCC region to further test whether the neural representations of group-based guilt were similar to those of personal guilt. We demonstrated that the classifier trained on the personal guilt conditions within the overlapped aMCC region can distinguish different levels of personal guilt in a leave-one-subject-out cross-validation test (accuracy = 68% ± 6%, $p < 0.001$; Fig. 6B). More importantly, the classifier can also discriminate the two group-based guilt conditions (*Out-group_ Commit vs. Out-group_ Observe*) with an accuracy of 71% ± 8%, $p = 0.01$ (Fig. 6B), demonstrating that the neural representations of personal guilt could discriminate those of group-based guilt.

## 4. Discussion

In this study, we revealed neurocognitive profiles of group-based guilt and explored its similarity with personal guilt. Specifically, shared responsibility for in-group transgressions is a crucial cognitive antecedent of group-based guilt just as personal guilt (Baumeister et al., 1994; Hoffman, 2001; Zahn-Waxler and Kochanska, 1990). Moreover, results of fMRI data suggested shared neuropsychological processes underlying these two forms of guilt. Our findings thus provide evidence for the Intergroup Emotion Theory of group-based emotion (Mackie et al., 2008; Smith and Mackie, 2015), which posits that the neurocognitive machinery for individual-level emotions are co-opted in the group context.

Previous research on personal guilt has demonstrated that the sense of responsibility in causing suffering to others is closely related to the experience of guilt (Hoffman, 2001; Zahn-Waxler and Kochanska, 1990). Similarly, the perceived moral responsibility is also an important antecedent of group-based guilt (Čehajić-Clancy et al., 2011; Iyer et al., 2004). As demonstrated in the present study, although the participants had no personal contribution to transgressions in the *Observe* conditions, they shared greater moral responsibility for transgressions committed by in-group partners than out-group partners, and this shared responsibility was positively associated with the guilt rating (*In-group_ Observe* > *Out-group_ Observe*). These results are consistent with the shared responsibility account of group-based guilt (Smiley, 2017; Tollefsen, 2006): that individuals who identify themselves with a group acting

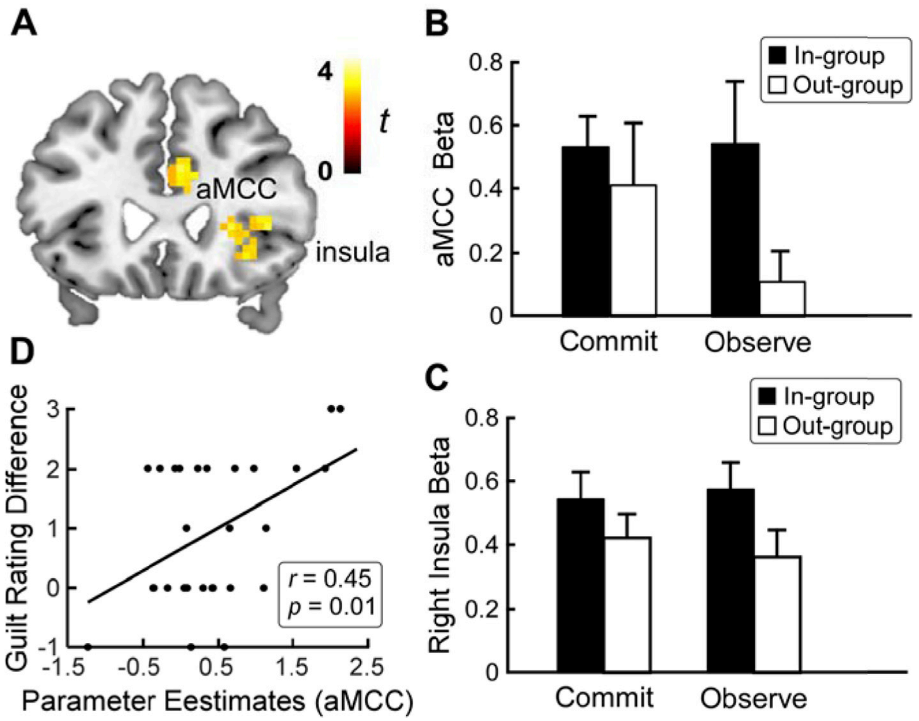


**Fig. 4. Brain activations related to group-based guilt.** Results of the contrast ‘*In-group_ Observe* > *Out-group_ Observe*’ are shown in yellow-to-red clusters (A). Statistical parametric map was displayed at $P < 0.005$ uncorrected at peak level with cluster size ≥46 voxels. Regional activation patterns (i.e., beta estimates) were extracted from aMCC (B) and rAI (C) regions-of-interest (27 voxels around the peak coordinates of aMCC (MNI coordinates = [6, 26, 28]) and rAI (MNI coordinates = [27, 20, −11])). (D) the parameter estimates of aMCC difference (Observe: *In-group_ Observe* > *Out-group_ Observe*) positively correlated with the post-experiment guilt rating difference (Observe: *In-group_ Observe* > *Out-group_ Observe*). Error bars represent standard errors of the means. Color should be used for Fig. 4 in print.

badly would feel jointly responsible for the bad behaviors of the group. According to social identity theory (Tajfel and Turner, 1986), when people categorize themselves as a member of a group, and internalize the group component into self-concept, the actions of other in-group members will have a direct influence on self-perception. Thus, this ‘vicarious’ moral responsibility for in-group transgressions may result from the shared group membership with the harm-doers (Tajfel and Turner, 1986; Wohl et al., 2006). In a broader sense, these results suggested that responsibility, inasmuch as it is relevant for appraisals of social emotions, need be construed more broadly to include the sense of responsibility an individual inherited from their group identities. More importantly, our study, perhaps for the first time, demonstrates that similarities between group-based and personal guilt go beyond self-reported emotional experience; the similarity is rooted in the neurobiology of emotional appraisal, lending support to the Intergroup Emotion Theory (Smith, 1993; Smith and Mackie, 2015). According to this theory, group members feel emotions in response to events affecting other in-group members as if those events were happening to themselves. An implication of this theory is that group-based emotions and personal emotions may share common neurocognitive substrates (Rydell et al., 2008). Our findings provide direct neural evidence for this hypothesis: aMCC responses to an in-group partners’ transgressions (relative to an out-group’s transgression) was positively associated with increased ratings of guilt, suggesting that aMCC is involved in experiencing group-based guilt.

Our findings shed new light on the role of aMCC in social-affective processing, extending its functional significance to inter-group processes (Lamm et al., 2011; Shackman et al., 2011a,b). Thus, the aMCC activations here for transgressions made by an in-group partner may reflect the vicarious ‘personal guilt’ elicited by the perceived moral responsibility for transgressions, and raising the possibility of a shared neuropsychological process underlying these two forms of guilt (Woo et al., 2014). This possibility was further corroborated by our multivariate analyses examining the similarity of neural representations of personal guilt and group-based guilt: the classifier trained on the personal guilt conditions could dissociate the group-based guilt conditions, suggesting that the group-based guilt could be developed on the basis of personal guilt.

An alternative way to understand the relationship between personal guilt and group-based guilt is worth noting, which takes a “family resemblance” perspective on the taxonomy of emotions (Shaver et al., 1987). Specifically, different species of guilt (e.g., personal guilt, group-based guilt, survivor guilt, etc.) are called “guilt” because they share a set of core cognitive-affective processes and phenomenology. Among them, personal guilt is the prototype of all guilt in that those core cognitive-affective processes are most clearly exemplified (Deigh, 1999; Morris, 1987). Although these two conceptualizations may have different ontological implications for the relationship between personal guilt and group-based guilt, testing such differences is beyond the scope of the current study. Future research with more sophisticated experimental design and more naturalistic materials (e.g., testimonies of real-world instances of group-based guilt) are needed to advance our understanding of the relationship between personal and group-based guilt.

Scenario-based imagination or recall of historical events that involve inter-group conflict have been widely used to induce group-based guilt in previous studies (Branscombe et al., 2004; Brown et al., 2008; Doosje et al., 1998; McGarty et al., 2005), which have served as a point of departure for further research on group-based guilt. However, the high ecological validity of this approach, although being a strength of this approach, often comes at the cost of a well-controlled experimental manipulation. For example, this approach typically involves comparisons between a scenario depicting inter-group harm event and a scenario depicting non-harmful neutral event, thereby confounding group-based guilt with the level of harm inflicted on the victims (Branscombe et al., 2004; Brown et al., 2008; Doosje et al., 1998; McGarty et al., 2005). Therefore, differences in self-reported guilt between conditions in these studies may be due to different levels of empathy or compassion for the victims’ harm. To avoid such a confound, in the current study, we induced group-based guilt not by manipulating levels of harm to the victims, but by manipulating the group membership of the transgressor (i.e., *In-group_ Observe* > *Out-group_ Observe*). Because the target and degree of harm is matched between the two conditions, any difference should be attributed to group membership (i.e., in-group/out-group), rather than to difference in vicarious pain, i.e. the spontaneous experience of pain when seeing another in pain (Corradi-Dell’Acqua et al., 2016; Krishnan et al., 2016; Lamm et al., 2011; Osborn and Derbyshire, 2010; Zaki et al., 2016).

Another limitation of the scenario-based method that many previous studies have adopted is that it confines the research on group-based guilt within populations who have a history of intergroup conflict (e.g., intergroup conflicts between Israeli and Palestinians; Wohl and Branscombe, 2008), which hinders the generalizability of the empirical evidence for conceptualizing group-based guilt as a universal psychological capacity. A strength of the interaction-based minimal group paradigm is that it extends the investigation of group-based guilt into populations without historical intergroup conflict. More importantly, the use of a minimal group paradigm provides a flexible experimental tool that could induce the intergroup emotion in a live and dynamic laboratory setting, thereby enabling a well-controlled laboratory study on group-based guilt to be conducted in natural circumstances. The investigation of the neuroscience of intergroup emotion is a relatively recent development within social psychology, but one that reflects a rapidly expanding interest in the interface between emotions and intergroup relations (Vollberg and Cikara, 2018). Following this trend, our interaction-based mini-group paradigm provides a flexible experimental tool to investigate the neural basis of group-based guilt.

According to an influential psychological account of guilt, empathy for the victim’s suffering is a core cognitive-affective component of guilt (Baumeister et al., 1994; Hoffman, 2001; Tangney et al., 2007). Empathy for others’ suffering, according to a quantitative meta-analysis of neuroimaging studies (Lamm et al., 2011), is consistently associated with cingulate cortex (including aMCC) and bilateral insula. Therefore, it is

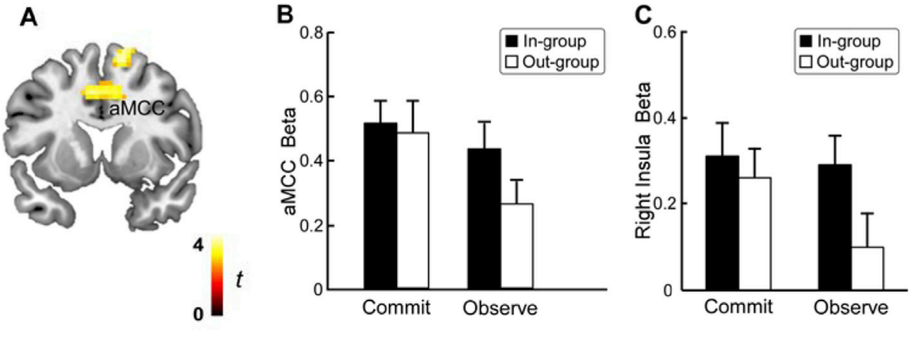


**Fig. 5. Brain activations related to personal guilt (‘*Out-group_ Commit* > *Out-group_ Observe*’).** (A). Results of the contrast ‘*Out-group_ Commit* > *Out-group_ Observe*’ are shown in yellow-to-red clusters. Threshold for display was $P < 0.005$ uncorrected at peak level with cluster size ≥46 voxels (B). Regional activation patterns (i.e., beta estimates). of aMCC was extracted from 27 voxels around the peak coordinates of aMCC (MNI coordinates = [12, 17, 40]). (C). Regional activation pattern of rAI was extracted from an independently defined region of interest (27 voxels around the peak coordinates of rAI reported in Yu et al., 2014; MNI coordinates = [36, 30, −8]). Error bars represent standard errors of the means. Color should be used for Fig. 5 in print.

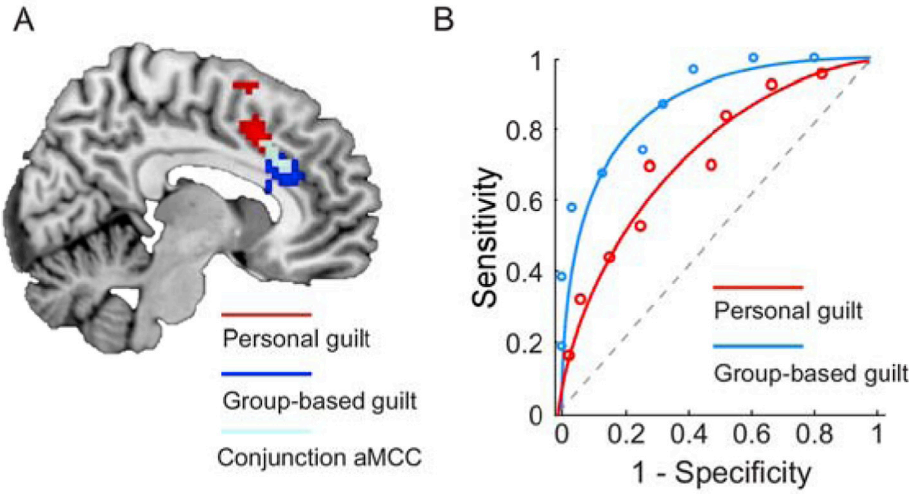


**Fig. 6. Results of multi-variate pattern analysis (MVPA). (A)** Common areas for personal guilt and group-based guilt in aMCC. Threshold for display was $P < .005$ uncorrected at peak level with cluster size $\geq 30$ voxels. **(B)** Receiver operating characteristic curves (ROC) for the two-choice forced-alternative accuracies. Red: Personal guilt; Blue: Group-based guilt. Color should be used for Fig. 6 in print.

difficult in the current study to dissociate the neural processes underlying empathy and guilt. Based on the data we have, we cannot rule out the possibility that the observed effects can be explained by empathy. In fact, the observed effects in self-reported guilt and allocation could be mediated by empathy. Future studies that include direct or implicit measures of empathy are needed to empirically test this hypothesis. That being said, compared with previous studies using scenario-based task, our design has already improved in terms of controlling confounding factors such as victims' harm across critical conditions.

It might be argued that if group-based guilt is elicited by observing other group members' transgressive behaviors, then it should rely on the "Theory of Mind (ToM)" processes. We did not observe stronger activations of the so-called ToM network (e.g., temporoparietal junction, dorsomedial prefrontal cortex, etc.) in the group-based guilt condition compared with the personal guilt condition. In theory, one can experience vicarious emotion in at least two ways: one can either 'inherit' the cognitive antecedents from others and generate their own emotions based on the shared cognitive antecedents (Lickel et al., 2005), or they can directly feel the emotion that the others express, without sharing or even knowing the cognitive antecedents leading to the others' emotion (e.g., it is not uncommon that we laugh when watching a group of strangers laughing, even when we do not know why the group is laughing in the first place). The second way, also known as emotional contagion (Hatfield et al., 1993), has been shown to activate ToM related brain areas (Nummenmaa et al., 2008; Melchers et al., 2015; Müller-Pinzler et al., 2016). In this study, the way we induced group-based guilt has made it conceptually closer to the first type of vicarious emotion: 1) the participants could not see their in-group members' face when they observed the latter causing pain to the victim, neither did they know whether their in-group members feel guilty at all, therefore it was unlikely that they discerned guilt from the in-group members and took on that feelings themselves; 2) our behavioral data did show that the participants 'inherited' their in-group members' responsibility (a core cognitive antecedent of guilt) in the transgression, which was predictive of their group-based guilt and reparation (Fig. 3D). Therefore, it is not surprising that the guilt-related rather than the ToM-related network played a more important role in this study. It is an empirical question as to whether and how watching a guilt display by an in-group member can contribute to group-based guilt over and above the vicarious processing of the in-group member's responsibility.

Note that it is beyond the scope of this study to empirically discriminate guilt from other related emotions, such as regret and shame. While it is still under debate in emotion science whether and how emotions should be categorized (Griffiths, 2004; Lindquist and Barrett, 2012; Satpute et al., 2016; Wager et al., 2015), we assume that emotions are characterized and differentiated, at least in part, by cognitive antecedents that give rise to them (Ellsworth and Smith, 1988; Frijda, 1993). As outlined above, the theory of interpersonal guilt that we adopted in this study underscores two critical antecedents of guilt, namely interpersonal harm and one's perceived contribution (though not necessarily moral responsibility) in causing the harm (Baumeister et al., 1994; Tangney and Dearing, 2003). We therefore define guilt as cognitive-affective processes triggered by the detection of the two antecedents in social encounters (see also Bastin et al., 2016; Cracco et al., 2015; Furukawa et al., 2019; Kédia et al., 2008; Koban et al., 2013; Leng et al., 2017; Seara-Cardoso et al., 2016; Yu et al., 2014). Phenomenologically, guilt typically involves a counterfactual desire that the violation that caused guilt had not happened in the first place (Baumeister et al., 1995). In that sense, guilt contains regret, or more precisely agent-regret, as its experiential component (Baron, 1988; Williams, 1976). When it comes to distinguishing guilt from shame, researchers have proposed two criteria (Ferguson et al., 2007; Tangney et al., 2007; Zhu et al., 2019): whether a violation is about a specific action (guilt) or about the subject as a person (shame), and whether a violation is private (guilt) or publicly known (shame). While these are by no mean clear-cut boundaries, the interpersonal harm situation adopted in the current study, i.e., unintentionally causing moderate physical discomfort to anonymous strangers with whom one would never meet or interact, arguably leans toward the guilt side of the guilt-shame dichotomy. To more clearly make that distinction, future work needs to incorporate more fine-grained experimental design to directly manipulate those relevant appraisals (e.g., action-vs. person-centered violation, publicity of the violation, etc.) combined with multivariate signature decoding approach (e.g., Ashar et al., 2017; Eisenbarth et al., 2016; Koban et al., 2019; Krishnan et al., 2016; Woo et al., 2014; Yu et al., 2019).

It is possible that group-based guilt and group-based shame are not mutually exclusive from each other in moral categories. Morally significant experiences can incur both guilt and shame, depending on which aspect of the truth is focused on. When focusing on the sympathy and pity for suffering of victim, it is more likely to feel guilty and adopt guilt-related behaviors, such as compensation or apology. However, when the focus switches to how this fact may threaten my self-value, such as "I" advocate freedom, equality, goodwill and so on, but the fact that "my" great-grandfather massacred civilians challenges my self-value, then "I" may tend to feel shame and adopt shame-related behaviors (e.g. hide, withdraw, denial). So different foci on the same fact can invoke different moral emotion and lead to different behavioral patterns. The increasing evidences from fMRI studies also suggested that guilt and shame are both distinct and coexist. For instance, the two emotions activated regions relate to the theory of mind (TPJ) and self-referential processing (ventral anterior cingulate cortex, vACC), but the further multivariate pattern analysis found that the neural patterns of the dorsal medial prefrontal cortex (dmPFC) and vACC could distinguish guilt and shame (Zhu et al., 2019). In summary, the boundary between guilt and shame is blur.

Although guilt feeling for in-group's transgressions may be a universal ability affiliation to distinct social groups, race and cultural values (individualism-collectivism) may also shape the extent to which one experiences the group-level guilt. Take the individualistic-collectivistic culture distinction as example, referring to the extent to which a culture weight the needs and values of self over those of a group, and prior neuroimaging studies have shown that the anterior rostral portion of the MPFC/ACC plays a unique role in the formation of collectivism (Wang et al., 2012). Usually, the East Asian individuals tend to be group oriented (Triandis, 1995) and emphasize collective identity, whereas the Euro-Americans are more self-oriented and emphasize private identity (Triandis, 1989). Thus, the experience of group-based guilt may be shaped by the individualism-collectivism. More specifically, East Asian individuals may experience stronger guilt for group wrongdoings than Euro-Americans, as individuals from the collectivistic cultures see themselves as fundamentally connected with others (Markus and Kitayama, 1991). Future cross-culture study should further explore culture differences on group-based guilt.

The fact that our participants were exclusively university students within a very specific age range limits to some extent the generalizability of our study. Future studies can improve the diversity and generalizability of our findings by recruiting participants who have directly experienced real-world social upheavals and compare their neurocognitive responses to inter-group conflict and transgressions. Nevertheless, taken together with previous studies on the same topic, our study has already contributed to the generalizability in two ways. First, to our knowledge, all the existing empirical studies on group-based (or collective) guilt are based on Western Caucasian populations (e.g. Baumeister et al., 1994; Čehajić-Clancy et al., 2011; Iyer et al., 2004). It is therefore not known whether group-based guilt is contingent on the social, political, religious and cultural conditions of Western society (e.g., Christianity, individualism, etc.). Therefore, like many other research topics in psychology, the generalizability of existing studies on group-based (collective) guilt is limited by the predominant (if not exclusive) reliance on the so-called W.E.I.R.D (Western, Educated, Industrialized, Rich and Democratic) population (Norenzayan et al., 2010). In this context, our study, by recruiting participants from an East Asian culture (i.e., Chinese) could help improving the diversity and generalizability of existing findings. Second, and more specific to the research on group-based guilt, our study has extended the findings of previous studies to a population that does not have any salient real-world inter-group conflict and atrocities. This demonstrates that group-based guilt is not contingent on having such a historical past and therefore could reflect a universal human emotion.

Our findings that group-based guilt has similar neural representation and action tendency as personal guilt have important societal implications. Even Hannah Arendt, who disputes the normative status of group-based guilt, agrees that crimes done by "our fathers or our people … may well make us pay for them" (Arendt, 2009). In today's increasingly polarizing and outrageous world, conflicts and transgressions between groups have become a serious threat to humanity and well-beings (Adam et al., 2014; Brady and Crockett, 2019; Crockett, 2017). In this study, reminding people their shared identity with a transgressor not only makes them feel responsible for the transgression, but also triggers a neural circuit that encodes personal guilt. As previous studies have shown, being in a (personal) guilt state makes people more supportive to policy that restores justice even at their personal cost (Gao et al., 2018; Wohl et al., 2019). Taken together, an important implication of our study is that emphasizing shared identity and responsibility within transgressor's group may facilitate policies aimed for restorative justice and inter-group reconciliation.

## 5. Conclusion

By combining fMRI with an interactive collective game, we provide behavioral and neural evidence demonstrating that group-based (collective) guilt and personal guilt share similar neurocognitive mechanisms. These findings provide neural evidence for the hypotheses that members of a group feel emotions in response to events affecting other in-group members as if those events affected them personally (Smith and Mackie, 2015). We highlight the crucial role of perceived responsibility in group-based guilt, which may help design interventions aimed at reducing unnecessary and excessive guilt inherited from past generations or other group members. Our findings shed light on the brain processes through which group membership is integrated into emotion appraisal, bridging gaps between research on emotion and inter-group relationship.

## Data and materials availability

The source data underlying Figs, 3A-D, 4D, Tables 1, 2, and Supplementary Fig. S1, Table S1, S2 are provided as a Source Data file. Contrast images are available on Neurovault (https://neurovault.org/my_collections/?q=), and code are available at GitHub (https://github.com/carolinelee0602/guilty-by-association). fMRI raw data are not publicized due to privacy concern.

## Declaration of competing interest

The authors declare that they have no competing interests.

## CRediT authorship contribution statement

**Zhiai Li:** Formal analysis, Writing - original draft. **Hongbo Yu:** Formal analysis, Writing - original draft. **Yongdi Zhou:** Writing - original draft. **Tobias Kalenscher:** Writing - original draft. **Xiaolin Zhou:** Writing - review & editing.

## Acknowledgements

This work was supported by National Natural Science Foundation of China (31630034) and by National Basic Research Program of China (973 Program: 2015CB856400).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116488.

## References

Aron, A., Aron, E.N., Smollan, D., 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. J. Personal. Soc. Psychol. 63, 596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Amodio, D.M., Devine, P.G., Harmon-Jones, E., 2007. A dynamic model of guilt: implications for motivation and self-regulation in the context of prejudice. Psychol. Sci. 18, 524–530. https://doi.org/10.1111%2Fj.1467-9280.2007.01933.x.

Arendt, H., 2009. Responsibility and Judgment. Schocken, New York.

Adam, D.I.K., Guillory, J.E., Hancock, J.T., 2014. Experimental evidence of massive-scale emotional contagion through social networks. Proc. Natl. Acad. Sci. U. S. A. 111, 8788–8790. https://doi.org/10.1073/pnas.1320040111.

Allpress, J.A., Brown, R., Giner-Sorolla, R., Deonna, J.A., Teroni, F., 2014. Two faces of group-based shame: moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. J. Personal. Soc. Psychol. 40, 1270–1284. https://doi.org/10.1177/0146167214540724.

Ashar, Y.K., Andrews-Hanna, J.R., Dimidjian, S., Wager, T.D., 2017. Empathic care and distress: predictive brain markers and dissociable brain systems. Neuron 94, 1263–1273. https://doi.org/10.1016/j.neuron.2017.05.014.

Bastin, C., Harrison, B.J., Davey, C.G., Moll, J., Whittle, S., 2016. Feelings of shame, embarrassment and guilt and their neural correlates: a systematic review. Neurosci. Biobehav. Rev. 71, 455–471. https://doi.org/10.1016/j.neubiorev.2016.09.019.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59, 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baron, M., 1988. Remorse and agent-regret. Midwest Stud. Philos. 13, 259–281. https://doi.org/10.1111/j.1475-4975.1988.tb00126.x.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Baumeister, R.F., Stillwell, A.M., Heatherton, T.F., 1994. Guilt: an interpersonal approach. Psychol. Bull. 115, 243–267. https://doi.org/10.1037/0033-2909.115.2.243.

Baumeister, R.F., Stillwell, A.M., Heatherton, T.F., 1995. Personal narratives about guilt: role in action control and interpersonal relationships. Basic Appl. Soc. Psychol. 17, 173–198. https://doi.org/10.1207/s15324834basp1701.

Branscombe, N.R., Slugoski, B., Kappen, D.M., 2004. Collective Guilt: what it Is and what it Is Not. Cambridge University Press., Cambridge.

Brown, R., González, R., Zagefka, H., Manzi, J., Čehajić, S., 2008. Nuestra culpa: collective guilt and shame as predictors of reparation for historical wrongdoing. J. Personal. Soc. Psychol. 94, 75–90. https://doi.org/10.1037/0022-3514.94.1.75.

Buckholtz, J.W., 2015. Social norms, self-control, and the value of antisocial behavior. CURR OPIN BEHAV SCI 3, 122–129. https://doi.org/10.1016/j.cobeha.2015.03.004.

Brady, W.J., Crockett, M.J., 2019. How effective is online outrage? Trends Cogn. Sci. 23, 79–80. https://doi.org/10.1016/j.tics.2018.11.004.

Cracco, E., Desmet, C., Brass, M., 2015. When your error becomes my error: anterior insula activation in response to observed errors is modulated by agency. Soc. Cogn. Affect. Neurosci. 11, 357–366. https://doi.org/10.1093/scan/nsv120.

Castano, E., Giner-Sorolla, R., 2006. Not quite human: infrahumanization in response to collective responsibility for intergroup killing. J. Personal. Soc. Psychol. 90, 804–818. https://doi.org/10.1037/0022-3514.90.5.804.

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, 371–378. https://doi.org/10.1016/j.neuroimage.2015.03.034.

Čehajić-Clancy, S., Effron, D.A., Halperin, E., Liberman, V., Ross, L.D., 2011. Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. J. Personal. Soc. Psychol. 101, 256–270. https://doi.org/10.1037/a0023936.

Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., Wager, T.D., 2015. A sensitive and specific neural signature for picture-induced negative affect. PLoS Biol. 13, e1002180 https://doi.org/10.1371/journal.pbio.1002180.

Chang, L.J., Smith, A., Dufwenberg, M., Sanfey, A.G., 2011. Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron 70, 560–572. https://doi.org/10.1016/j.neuron.2011.02.056.

Crockett, M.J., 2017. Moral outrage in the digital age. Nat. Hum. Behav. 1, 769–771. https://doi.org/10.1038/s41562-017-0213-3.

Crockett, M.J., Siegel, J.Z., Kurth-Nelson, Z., Dayan, P., Dolan, R.J., 2017. Moral transgressions corrupt neural representations of value. Nat. Neurosci. 20, 879–885. https://doi.org/10.1038/nn.4557.

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., Singer, T., 2016. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. Nat. Commun. 7, 10904. https://doi.org/10.1038/ncomms10904.

Coutanche, M.N., 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? COGN AFFECT BEHAV NE 13, 667–673. https://doi.org/10.3758/s13415-013-0186-2.

Diefendorff, J.M., Richard, E.M., 2003. Antecedents and consequences of emotional display rule perceptions. Pol. J. Appl. Psychol. 88, 284–294. https://doi.org/10.1037/0021-9010.88.2.284.

Doosje, B., Branscombe, N.R., Spears, R., Manstead, A.S., 1998. Guilty by association: when one's group has a negative history. J. Personal. Soc. Psychol. 75, 872–886. https://doi.org/10.1037//0022-3514.75.4.872.

Doosje, B., Branscombe, N.R., Spears, R., Manstead, A.S., 2004. Consequences of national ingroup identification for responses to immoral historical events. In: Branscombe, N.B., Branscombe, N.R., Doosje, B. (Eds.), Collective Guilt: International Perspectives. Cambridge University Press., Cambridge, pp. 95–111.

Dunham, Y., 2018. Mere membership. Trends Cogn. Sci. 22, 780–793. https://doi.org/10.1016/j.tics.2018.06.004.

Deigh, J., 1999. All Kinds of Guilt. Law and Philosophy, vol. 18, pp. 313–325. https://doi.org/10.1023/A:1006380226393.

Eisenbarth, H., Chang, L.J., Wager, T.D., 2016. Multivariate brain prediction of heart rate and skin conductance responses to social threat. J. Neurosci. 36, 11987–11998. https://doi.org/10.1523/JNEUROSCI.3672-15.2016.

Ellsworth, P.C., Smith, C.A., 1988. From appraisal to emotion: differences among unpleasant feelings. Motiv. Emot. 12, 271–302. https://repository.law.umich.edu/articles/1669.

Falk, C.F., Heine, S.J., Takemura, K., 2014. Cultural variation in the minimal group effect. J. Cross Cult. Psychol. 45, 265–281. https://doi.org/10.1177/0022022113492892. https://search.crossref.org/?q=Cultural+variation+in+the+minimal+group+effect.

Ferguson, T.J., Brugman, D., White, J., Eyre, H.L., 2007. Shame and guilt as morally warranted experiences. In: Tracy, J.L., Robins, R.W., Tangney, J.P. (Eds.), The Self-Conscious Emotions: Theory and Research. Guilford Press., New York, pp. 330–348.

Ferguson, M.A., Branscombe, N.R., 2014. The social psychology of collective guilt. In: Salmela, M., Scheve, C.V. (Eds.), Collective Emotions: Perspectives from Psychology, Philosophy, and Sociology. Oxford UP., Oxford, pp. 251–265.

Frijda, N.H., 1993. The place of appraisal in emotion. Cognit. Emot. 7, 357–387.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer series in statistics, New York.

Friston, K.J., Penny, W.D., Glaser, D.E., 2005. Conjunction revisited. Neuroimage 25, 661–667. https://doi.org/10.1016/j.neuroimage.2005.01.013.

Furukawa, Y., Nakashima, K.I., Tsukawaki, R., Morinaga, Y., 2019. Guilt as a signal informing us of a threat to our morality. Curr. Psychol. 1–11 https://doi.org/10.1007/s12144-019-0144-4.

Griffiths, P.E., 2004. Emotions as natural and normative kinds. Philos. Sci. 71, 901–911. https://doi.org/10.1086/425944.

Griswold, C., 2007. Forgiveness: A Philosophical Exploration. Cambridge University Press, Cambridge.

Gao, X., Yu, H., Sáez, I., Blue, P.R., Zhu, L., Hsu, M., Zhou, X., 2018. Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. Proc. Natl. Acad. Sci. 115, E7680–E7689. https://doi.org/10.1073/pnas.1802523115.

Goldin, P.R., McRae, K., Ramel, W., Gross, J.J., 2008. The neural bases of emotion regulation: reappraisal and suppression of negative emotion. Biol. Psychiatry 63, 577–586. https://doi.org/10.1016/j.biopsych.2007.05.031.

Hatfield, E., Cacioppo, J.T., Rapson, R.L., 1993. Emotional contagion. Curr. Dir. Psychol. Sci. 2, 96–100. https://doi.org/10.1111%2F1467-8721.ep10770953.

Halperin, E., Schori-Eyal, N., 2019. Moral Emotions in Political Decision Making. Oxford University Press, Oxford.

Hoffman, M.L., 2001. Empathy and Moral Development: Implications for Caring and Justice. Cambridge University Press, Cambridge.

Immordino-Yang, M.H., McColl, A., Damasio, H., Damasio, A., 2009. Neural correlates of admiration and compassion. Proc. Natl. Acad. Sci. 106, 8021–8026. https://doi.org/10.1073/pnas.0810363106.

Iyer, A., Leach, C.W., Pedersen, A., 2004. Racial wrongs and restitutions: the role of guilt and other group-based emotions. In: Fine, M., Weis, L., Pruitt, L.P., Burns, A. (Eds.), Off White: Readings on Power, Privilege, and Resistance. Routledge., London, pp. 345–361.

Izard, C.E., 1991. The Psychology of Emotions. Plenum, New York.

Inui, K., Tran, T.D., Qiu, Y., Wang, X., Hoshiyama, M., Kakigi, R., 2002. Pain-related magnetic fields evoked by intra-epidermal electrical stimulation in humans. Clin. Neurophysiol. 113, 298–304. https://doi.org/10.1016/S1388-2457(01)00734-9.

Johnson, D.W., Johnson, F.P., 1991. Joining Together: Group Theory and Group Skills. Prentice Hall, Englewood Cliffs, NJ.

Kédia, G., Berthoz, S., Wessa, M., Hilton, D., Martinot, J.L., 2008. An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. J. Cogn. Neurosci. 20, 1788–1798. https://doi.org/10.1162/jocn.2008.20070.

Krishnan, A., Woo, C.-W., Chang, L.J., Ruzic, L., Gu, X., López-Solà, M., Wager, T.D., 2016. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. eLife 5 (2016). https://doi.org/10.7554/eLife.15166.

Ketelaar, T., Tung Au, W., 2003. The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: an affect-as-information interpretation of the role of emotion in social interaction. Cognit. Emot. 17, 429–453. https://doi.org/10.1080/02699930143000662.

Koban, L., Corradi-Dell'Acqua, C., Vuilleumier, P., 2013. Integration of error agency and representation of others' pain in the anterior insula. J. Cogn. Neurosci. 25, 258–272. https://doi.org/10.1162/jocn_a_00324.

Koban, L., Jepma, M., López-Solà, M., Wager, T.D., 2019. Different brain networks mediate the effects of social and conditioned expectations on pain. Nat. Commun. 10, 1–13. https://doi.org/10.1038/s41467-019-11934-y.

Lickel, B., Steele, R.R., Schmader, T., 2011. Group-based shame and guilt: emerging directions in research. Soc Personal Psychol Compass 5, 153–163. https://doi.org/10.1111/j.1751-9004.2010.00340.x.

Lickel, B., Schmader, T., Curtis, M., Scarnier, M., Ames, D.R., 2005. Vicarious shame and guilt. Group Process. Intergr. Relat. 8, 145–157. https://doi.org/10.1177%2F1368430205051064.

Leng, B., Wang, X., Cao, B., Li, F., 2017. Frontal negativity: an electrophysiological index of interpersonal guilt. Soc. Neurosci. 12, 649–660. https://doi.org/10.1080/17470919.2016.1223749.

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, 2492–2502. https://doi.org/10.1016/j.neuroimage.2010.10.014.

Lieberman, M.D., Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc. Cogn. Affect. Neurosci. 4, 423–428. https://doi.org/10.1093/scan/nsp052.

Lindquist, K.A., Barrett, L.F., 2012. A functional architecture of the human brain: emerging insights from the science of emotion. Trends Cogn. Sci. 16, 533–540. https://doi.org/10.1016/j.tics.2012.09.005.

Mackie, D.M., Smith, E.R., Ray, D.G., 2008. Intergroup emotions and intergroup relations. Soc Personal Psychol Compass 2, 1866–1880. https://doi.org/10.1111/j.1751-9004.2008.00130.x.

Matsumoto, D., 1993. Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. Motiv. Emot. 17, 107–123. https://link.springer.com/article/10.1007/BF00995188.

McGarty, C., Pedersen, A., Wayne Leach, C., Mansell, T., Waller, J., Bliuc, A.M., 2005. Group-based guilt as a predictor of commitment to apology. Br. J. Psychol. 44, 659–680. https://doi.org/10.1348/014466604X18974.

McRae, K., Hughes, B., Chopra, S., Gabrieli, J.D.E., Gross, J.J., Ochsner, K.N., 2010. The neural bases of distraction and reappraisal. J. Cogn. Neurosci. 22, 248–262. https://doi.org/10.1162/jocn.2009.21243.

Morris, H., 1987. Nonmoral guilt. In: Schoeman, F. (Ed.), Responsibility, Character, and the Emotions: New Essays in Moral Psychology. Cambridge University Press., New York, pp. 220–240.

Müller-Pinzler, L., Krach, S., Krämer, U.M., Paulus, F.M., 2016. The social neuroscience of interpersonal emotions. Curr Top Behav Neurosci 30, 241–256. https://doi.org/10.1007/7854_2016_437.

Mesquita, B., Boiger, M., De Leersnyder, J., 2016. The cultural construction of emotions. Curr Opin Psychol 8, 31–36. https://doi.org/10.1016/j.copsyc.2015.09.015.

Markus, H.R., Kitayama, S., 1991. Culture and the self: implications for cognition, emotion, and motivation. Psychol. Rev. 98, 224. https://psycnet.apa.org/doi/10.1037/0033-295X.98.2.224.

Melchers, M., Markett, S., Montag, C., Trautner, P., Weber, B., Lachmann, B., et al., 2015. Reality TV and vicarious embarrassment: an fMRI study. Neuroimage 109, 109–117. https://10.1016/j.neuroimage.2015.01.022.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005.

Norenzayan, A., Henrich, J., Heine, S.J., 2010. Most people are not WEIRD. Nature 466. https://doi.org/10.1038/466029a, 29-29.

Nummenmaa, L., Hirvonen, J., Parkkola, R., Hietanen, J.K., 2008. Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. Neuroimage 43, 571–580. https://doi.org/10.1016/j.neuroimage.2008.08.014.

Osborn, J., Derbyshire, S.W.G., 2010. Pain sensation evoked by observing injury in others. Pain 148, 268–274. https://doi.org/10.1016/j.pain.2009.11.007.

Otten, S., 2016. The Minimal Group Paradigm and its maximal impact in research on social categorization. Curr Opin Psychol 11, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010.

Perron Tollefsen, D., 2003. Participant reactive attitudes and collective responsibility. Philos. Explor. 6, 218–234. https://doi.org/10.1080/10002003098538751.

Powell, A.A., Branscombe, N.R., Schmitt, M.T., 2005. Inequality as ingroup privilege or outgroup disadvantage: the impact of group focus on collective guilt and interracial attitudes. Personal. Soc. Psychol. Bull. 31, 508–521. https://doi.org/10.1177/0146167204271713. https://search.crossref.org/?q=Inequality+as+ingroup+privilege+or+outgroup+disadvantage%3A+the+impact+of+group+focus+on+collective+guilt+and+interracial+attitudes.

Price, C.J., Friston, K.J., 1997. Cognitive conjunction: a new approach to brain activation experiments. Neuroimage 5, 261–270. https://doi.org/10.1006/nimg.1997.0269.

Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav. Res. Methods 40, 879–891. https://link.springer.com/article/10.3758/BRM.40.3.879.

Rydell, R.J., Mackie, D.M., Maitner, A.T., Claypool, H.M., Ryan, M.J., Smith, E.R., 2008. Arousal, processing, and risk taking: consequences of intergroup anger. Personal. Soc. Psychol. Bull. 34, 1141–1152. https://doi.org/10.1177/0146167208319694. In: https://search.crossref.org/?q=Arousal%2C+processing%2C+and+risk+taking%3A+consequences+of+intergroup+anger.

Radke, S., De Lange, F.P., Ullsperger, M., De Bruijn, E.R.A., 2011. Mistakes that affect others: an fMRI study on processing of own errors in a social context. Exp. Brain Res. 211, 405–413. https://doi.org/10.1007/s00221-011-2677-0.

Satpute, A.B., Nook, E.C., Narayanan, S., Shu, J., Weber, J., Ochsner, K.N., 2016. Emotions in "black and white" or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. Psychol. Sci. 27, 1428–1442. https://doi.org/10.1177/0956797616661555. https://search.crossref.org/?q=Emotions+in+%E2%80%9Cblack+and+white%E2%80%9D+or+shades+of+gray%3F+How+we+think+about+emotion+shapes+our+perception+and+neural+representation+of+emotion.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game. Science 300, 1755–1758. https://doi.org/10.1126/science.1082976.

Seara-Cardoso, A., Sebastian, C.L., McCrory, E., Foulkes, L., Buon, M., Roiser, J.P., Viding, E., 2016. Anticipation of guilt for everyday moral transgressions: the role of the anterior insula and the influence of interpersonal psychopathic traits. Sci. Rep. 6, 36273. https://doi.org/10.1038/srep36273.

Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011a. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12, 154–167. https://doi.org/10.1038/nrn2994.

Smiley, M., 2017. Collective intentions and collective moral responsibility. In: Jankovic, M., Ludwig, K. (Eds.), The Routledge Handbook of Collective Intentionality. Routledge., London, pp. 316–326.

Smith, E.R., 1993. Social identity and social emotions: toward new conceptualizations of prejudice. In: Mackie, D.M., Hamilton, D.L. (Eds.), Affect, Cognition and Stereotyping. Academic Press., Cambridge, pp. 297–315.

Smith, A.T., Kosillo, P., Williams, A.L., 2011. The confounding effect of response amplitude on MVPA performance measures. Neuroimage 56, 525–530. https://doi.org/10.1016/j.neuroimage.2010.05.079.

Smith, E.R., Mackie, D.M., 2015. Dynamics of group-based emotions: insights from intergroup emotions theory. EMR 7, 349–354. https://doi.org/10.1177/1754073915590614. https://search.crossref.org/?q=Dynamics+of+group-based+emotions%3A+insights+from+intergroup+emotions+theory.

Schoeman, F.D., 1987. Responsibility, Character, and the Emotions: New Essays in Moral Psychology. Cambridge University Press, New York.

Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011b. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12, 154–167. https://doi.org/10.1038/nrn2994.

Shaver, P., Schwartz, J., Kirson, D., O'Connor, C., 1987. Emotion knowledge: further exploration of a prototype Approach. J. Personal. Soc. Psychol. 52, 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061.

Tangney, J.P., Dearing, R.L., 2003. Shame and Guilt. Guilford Press, New York.

Tangney, J.P., Stuewig, J., Mashek, D.J., 2007. Moral emotions and moral behavior. Annu. Rev. Psychol. 58, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145.

Triandis, H.C., 1989. The self and social behavior in differing cultural contexts. Psychol. Rev. 96, 506–520. https://doi.org/10.1037/0033-295X.96.3.506.

Triandis, H.C., 1995. The importance of contexts in studies of diversity. In: Jackson, S.E., Ruderman, M.N. (Eds.), Diversity in Work Teams: Research Paradigms for a Changing Workplace. DC. American Psychological Association., Washington, pp. 225–233.

Taylor, G., 1985. Shame, Pride, and Guilt: Emotions of Self-Assessment. Oxford University Press, Oxford.

Tajfel, H., Turner, J.C., 1986. An integrative theory of intergroup relations. In: Worchel, S., Austin, W.G. (Eds.), Psychology of Intergroup Relations. Nelson-Hall., Chicago, pp. 7–24.

Tollefsen, D., 2006. The rationality of collective guilt. Midwest Stud. Philos. 30, 222–239. https://doi.org/10.1111/j.1475-4975.2006.00136.x.

Vollberg, M.C., Cikara, M., 2018. The neuroscience of intergroup emotion. Curr Opin Psychol 24, 48–52. https://doi.org/10.1016/j.copsyc.2018.05.003.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.-W., Kross, E., 2013. An fMRI-based neurologic signature of physical pain. N. Engl. J. Med. 368, 1388–1397. https://doi.org/10.1056/NEJMoa1204471.

Wager, T.D., Kang, J., Johnson, T.D., Nichols, T.E., Satpute, A.B., Barrett, L.F., 2015. A Bayesian model of category-specific emotional brain responses. PLoS Comput. Biol. 11, e1004066 https://doi.org/10.1371/journal.pcbi.1004066.

Williams, B., 1976. Moral luck. In: Williams, B. (Ed.), Proceedings of the Aristotelian Society, Supplementary Volume L. Cambridge University Press., Cambridge, pp. 115–135.

Wohl, M.J., Branscombe, N.R., Klar, Y., 2006. Collective guilt: emotional reactions when one's group has done wrong or been wronged. Eur. Rev. Soc. Psychol. 17, 1–37. https://doi.org/10.1080/10463280600574815.

Wohl, M.J., Branscombe, N.R., 2008. Remembering historical victimization: collective guilt for current ingroup transgressions. J. Personal. Soc. Psychol. 94, 988–1006. https://doi.org/10.1037/0022-3514.94.6.988.

Wohl, M.J., Tabri, N., Hollingshead, S.J., Dupuis, D.R., Caouette, J., 2019. Empathetic collective angst predicts perpetrator group members' support for the empowerment of the victimized group. J. Personal. Soc. Psychol. 117, 1083–1104. https://doi.org/10.1037/pspi0000176.

Wang, G., Mao, L., Ma, Y., Yang, X., Cao, J., Liu, X., Han, S., 2012. Neural representations of close others in collectivistic brains. Soc. Cogn. Affect. Neurosci. 7, 222–229. https://doi.org/10.1093/scan/nsr002.

Woo, C.-W., Koban, L., Kross, E., Lindquist, M.A., Banich, M.T., Ruzic, L., Wager, T.D., 2014. Separate neural representations for physical pain and social rejection. Nat. Commun. 5, 5380. https://doi.org/10.1038/ncomms6380.

Yu, H., Hu, J., Hu, L., Zhou, X., 2014. The voice of conscience: neural bases of interpersonal guilt and compensation. Soc. Cogn. Affect. Neurosci. 9, 1150–1158. https://doi.org/10.1093/scan/nst090.

Yu, H., Koban, L., Chang, L.J., Wagner, U., Krishnan, A., Vuilleumier, P., et al., 2019. A generalizable multivariate brain pattern for interpersonal guilt. Cerebr. Cortex. https://doi.org/10.1093/cercor/bhz326.

Zaki, J., Ochsner, K.N., Hanelin, J., Wager, T.D., Mackey, S.C., 2007. Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. Soc. Neurosci. 2, 276–291. https://doi.org/10.1080/17470910701401973.

Zaki, J., Wager, T.D., Singer, T., Keysers, C., Gazzola, V., 2016. The anatomy of suffering: understanding the relationship between nociceptive and empathic pain. Trends Cogn. Sci. 20, 249–259. https://doi.org/10.1016/j.tics.2016.02.003.

Zahn-Waxler, C., Kochanska, G., 1990. The origins of guilt. In: Thompson, R.A. (Ed.), Current Theory and Research in Motivation, Nebraska Symposium on Motivation, 1988: Socioemotional Development, vol. 36. University of Nebraska Press., Lincoln, pp. 183–258.

Zhu, R., Feng, C., Zhang, S., Mai, X., Liu, C., 2019. Differentiating guilt and shame in an interpersonal context with univariate activation and multivariate pattern analyses. Neuroimage 186, 476–486. https://doi.org/10.1016/j.neuroimage.2018.11.012.