

The Neural Mechanism of the Social Framing Effect: Evidence from fMRI and tDCS Studies

Jie Liu,^{1,2*} Ruolei Gu,^{3,4*} Chong Liao,¹ Juanzhi Lu,¹ Yuxing Fang,^{5,6} Pengfei Xu,^{1,2} Yue-jia Luo,^{1,2} and Fang Cui^{1,2}

¹School of Psychology, Shenzhen University, Shenzhen, China, 518060, ²Center for Brain Disorders and Cognitive Neuroscience, Shenzhen, China, 518060, ³CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China, 100101, ⁴Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, 100049, ⁵National Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, 100875, and ⁶Department of Psychology, University of Cambridge, Cambridge, United Kingdom, CB2 3EB

As an important cognitive bias, the framing effect shows that our decision preferences are sensitive to the verbal description (i.e., frame) of options. This study focuses on the neural underpinnings of the social framing effect, which is based on decision-making regarding other people. A novel paradigm was used in which participants made a trade-off between economic benefits and the feelings of others. This decision was described as either a “harm” to, or “not helping,” other persons in two conditions (Harm frame vs Help frame). Both human males and females were recruited. Participants behaved more prosocially for Harm frame compared with Help frame, resulting in a significant social framing effect. Using functional magnetic resonance imaging, Experiment 1 showed that the social framing effect was associated with stronger activation in the temporoparietal junction (TPJ), especially its right part. The functional connectivity between the right TPJ (rTPJ) and medial prefrontal cortex predicted the social framing effect on the group level. In Experiment 2, we used transcranial direct current stimulation to modulate the activity of the rTPJ and found that the social framing effect became more prominent under anodal (excitatory) stimulation, while the nonsocial framing effect elicited by the economic gain/loss gambling frame remained unaffected. The rTPJ results might be associated with moral conflicts modulated by the social consequences of an action or different levels of mentalizing with others under different frame conditions, but alternative interpretations are also worth noting. These findings could help elucidate the psychological mechanisms of the social framing effect.

Key words: decision-making; functional magnetic resonance imaging; multivoxel pattern analysis; right temporoparietal junction; social framing effect; transcranial direct current stimulation

Significance Statement

Previous studies have suggested that the framing effect is generated from an interaction between the amygdala and anterior cingulate cortex. This opinion, however, is based on findings from nonsocial framing tasks. Recent research has highlighted the importance of distinguishing between the social and nonsocial framing effects. The current study focuses on the social framing effect and finds out that the temporoparietal junction and its functional connectivity with the medial prefrontal cortex play a significant role. Additionally, modulating the activity of this region leads to changes in social (but not nonsocial) framing effect. Broadly speaking, these findings help understand the difference in neural mechanisms between social and nonsocial decision-making. Meanwhile, they might be illuminating to promote helping behavior in society.

Received June 13, 2019; revised Mar. 18, 2020; accepted Mar. 19, 2020.

Author contributions: J. Liu, C.L., and J. Lu performed research; J. Liu, Y.F., and P.X. analyzed data; J. Liu, R.G., and F.C. wrote the first draft of the paper; J. Liu, R.G., Y.L., and F.C. edited the paper; R.G. and F.C. designed research; Y.L. contributed unpublished reagents/analytic tools.

This work was supported by National Natural Science Foundation of China 31871109, 31571124, and 31900779, Youth Innovation Promotion Association CAS 2019088, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions 2019SHIBS0003, and Major Program of the Chinese National Social Science Foundation 17ZDA324. All procedures performed in this study were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The local ethics committee approved the experimental protocol.

The authors declare no competing financial interests.

*J. Liu and R.G. contributed equally to this work.

Correspondence should be addressed to Fang Cui at cuifang0826@gmail.com.

<https://doi.org/10.1523/JNEUROSCI.1385-19.2020>

Copyright © 2020 the authors

Introduction

In humans, the preference of different options could change dramatically in response to two objectively, but not verbally, equivalent statements (i.e., “frames”) of an identical problem (Rabin, 1998). This “framing effect” phenomenon deviates from standard economic assumptions (e.g., the axiom of descriptive invariance) and thus has become a hot research topic (Tversky and Kahneman, 1981). Framing research is not a homogeneous endeavor; rather, researchers have proposed various taxonomies (Kühberger, 1998; Levin et al., 1998). Notably, the social framing effect could be distinguished from the nonsocial one according to whether a social dilemma between the self and others is

involved in the current scenario. In our opinion, a social framing effect manifests when changing the description of a social dilemma (or a specific social component in this dilemma) significantly modulates a decision-maker's preference to different options. The psychological mechanisms of the nonsocial framing effect and that of the social one are essentially different (Ellingsen et al., 2012). When making nonsocial decisions (e.g., gambling), people focus on maximizing their individual utility, and decision frames affect their judgment of which option is more beneficial or less risky. In contrast, when making social decisions, people take the consequence of their actions to others into account, and decision frames affect people's preference through the influence of other-regarding concerns and social norms, such as fairness, altruism, and reciprocity (Van Dijk and Wilke, 2000). For instance, labeling a task as the "Community Game" rather than the "Wall Street Game" increased participants' cooperation rate (Liberman et al., 2004).

Studies on the neural underpinnings of the framing effect have proliferated in the last decade. To our knowledge, these studies have overwhelmingly focused on nonsocial framing, especially a gain/loss gamble framing task. Using this task, De Martino et al. (2006) first revealed that the amygdala activated more strongly when participant decisions were consistent with the nonsocial framing effect, whereas the anterior cingulate cortex (ACC) activated more strongly when decisions ran counter to this effect (see also Roiser et al., 2009; Guitart-Masip et al., 2010; Xu et al., 2013; Li et al., 2017; Wang et al., 2017). According to De Martino et al. (2006), these neural activity patterns modulated approach/avoidance tendencies on the individual level. The current study is interested in the neural underpinnings of the social framing effect. In our opinion, this effect would recruit specific regions that play important roles in social-cognitive processes and in decision-making (Parkinson and Wheatley, 2015; Tavares et al., 2015). According to the literature, the candidate regions include (but are not limited to) the temporoparietal junction (TPJ), inferior frontal gyrus, posterior cingulate cortex, and superior temporal sulcus (Sanfey, 2007; Rilling and Sanfey, 2011).

With two experiments, this study aimed to redress the paucity of knowledge using a paradigm in which the social implication of an option was emphasized or obscured. Specifically, we asked participants to make a trade-off between economic benefits and the feelings of others; when participants showed stronger preference to income maximization, the probability for their partners to receive a painful electrical shock would increase proportionally. This decision was described as either a "harm" to, or simply "not helping," other persons in two frame conditions. In this way, we hoped to distinguish between the influence of social processes and the approach/avoidance emotions elicited by economic gains or losses. Considering that avoid harming others is a strong, universal motive among human beings (Crockett et al., 2014), we hypothesized that participants would make more costly helping decisions in the Harm frame compared with the Help frame condition, resulting in a significant social framing effect. In Experiment 1, we explored with functional magnetic resonance imaging (fMRI) whether decisions consistent with the social framing effect would activate specific brain areas. Then in Experiment 2, we examined with transcranial direct current stimulation (tDCS) if temporarily modulating the activity of key regions found in Experiment 1 would lead to changes in the social framing effect.

Experiment 1 (fMRI)

Materials and methods

Participants

Thirty right-handed participants were recruited from Shenzhen University to join in the fMRI experiment. Participants were screened for a history of neurologic disorders, brain injury, and developmental disabilities. All had a normal or corrected-to-normal vision. Three of them who had excessive head movements $>2^\circ$ in rotation or >2 mm in translation during the scanning were excluded, leaving 27 participants in the final sample (13 women, age: 21.0 ± 1.8 years, mean \pm SD). The study was conducted according to the ethical guidelines and principles of the Declaration of Helsinki and was approved by the Medical Ethical Committee of Shenzhen University Medical School. Informed consent was obtained from all participants after they fully understood the procedures.

Experimental design and statistical analysis

Upon arrival at the laboratory, each participant was introduced to another participant (who was actually a confederate of the same sex). The two of them drew lots to decide their roles in the formal task, which were manipulated such that the real participant would always play as a "decider" while the confederate would play as a "victim." The real participant was then led to the fMRI environment; this participant was told that the victim would be in another room and that they would complete the task simultaneously via web connections.

Our paradigm was generally adapted from the "Approach-Avoidance Conflict" task, which has been proven to be suitable for fMRI research (Aupperle et al., 2011, 2015). In addition, the idea of providing a dilemma between subjective cost and physical pain for other people was inspired by Crockett et al. (2014), which pointed out that "an aversion to the suffering of others is a powerful motivator for humans." Before the scanning day, all participants were asked to submit a digital photograph that was expected to be used if they were assigned to the role of "victim" in the experiment. In reality, no participant would serve the role of "victim," and their photographs were not used for any purpose. The experiment applied a one-factor (social frames: Help frame vs Harm frame) within-subject design. In each trial, a photograph of the confederate ("victim") would be presented for 2 s to indicate the beginning of that trial. Then, each participant was shown a runway with pictures and texts on its left and right ends to represent two possible outcomes. In the "Help frame" condition, one outcome was "help the other person to avoid receiving a painful shock and subtract 5 RMB from your own payment," whereas the other outcome was "do not help the other person to avoid receiving a shock and keep all your payment." In the "Harm frame" condition, one outcome was "harm the other person by administering a painful shock and keep all your payment," whereas the other outcome was "do not harm the other person by administering a shock and subtract 5 RMB from your own payment" (Fig. 1A).

In both frame conditions, each participant essentially faced the same "costly helping" decision. The only difference between the two conditions was how the outcomes were framed. Specifically, the outcome "the participant keeps all money and the victim receives a painful shock" was described as not helping others in the "Help frame" condition but as intentionally harming others in the "Harm frame" condition, whereas the outcome "the participant loses 5 RMB and the victim receives no shock" was described as helping others in the "Help frame" condition but as avoid harming others in the "Harm frame" condition.

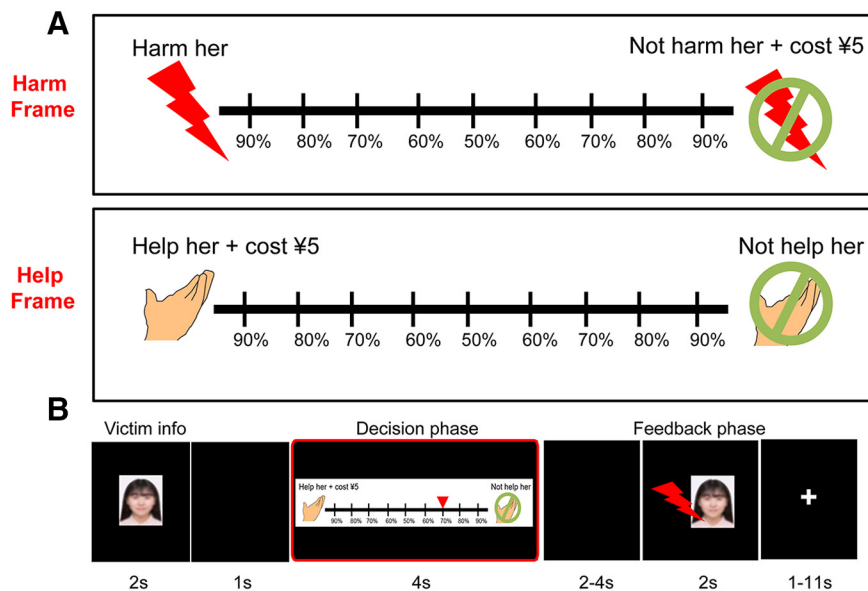


Figure 1. Experimental design. **A**, The “Help” and “Harm” frames. **B**, Structure of a trial. The target event for fMRI data analysis was the decision phase (red rectangle).

Each participant was asked to position an avatar on the runway to indicate his/her relative preference between the two potential outcomes. The final placement of the avatar would determine the occurrence probability of the two possible outcomes. For example, if the avatar was placed at the midpoint of the runway, the probability of each outcome was 50%; if the avatar was placed all the way to one side, the probability of its neighboring outcome was 90% and that of its farside outcome was 10%. The resolution of the slider was 10%, and the outcome position on either side of the slider was counterbalanced across trials. The participants were informed in the task instruction phase that they can influence the probabilities of the two outcomes by moving the avatar; otherwise, if they do not want to influence the outcome probability, then they can leave the avatar to the 50% point. The starting point of the avatar was randomized between trials, and the participants had 4 s to respond (i.e., decision phase). To respond, each participant pressed one of two preassigned buttons on an MRI-compatible button-box to move the avatar and pressed a third button to confirm its final place. After a 2–4 s waiting time, the feedback would be presented for 2 s. The interstimulus intervals were set to 1–11 s (Fig. 1B).

The feedback was generated by the computer according to the probability chosen by each participant, which serves as a reinforcing cue to inform the participants that their decisions were valid. Before the experiment, the participants were instructed that, throughout the task, only the feedback of 10 random trials would be executed. To avoid this feedback from influencing participant decisions in the following trials, they were not informed on which trial the feedback was actually executed. Participants were told that they would be informed about the outcomes of the executed trials after the scanning phase. These 10 trials would be randomly selected by the computer; thus, the best strategy for each participant was to treat each trial equally.

The experiment consisted of two runs of 40 trials lasting ~30 min. Each run included 20 trials of each frame condition (Help/Harm) that were pseudo-randomized. Thus, each condition contained 40 trials in total. Before the scanning, participants were familiarized with the task with a practice block consisting of eight trials. After the experiment, the participants received remuneration according to their task performance (120–150 Chinese

RMB, ~20 U.S. dollars) and a debriefing was given by the experimenter.

Neuroimaging data acquisition and preprocessing

We used a Siemens TrioTim 3.0T MRI machine for data acquisition. Functional volumes were acquired using multiple slice T2-weighted echo planar imaging sequences with the following parameters: repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, field of view = 224 × 224 mm², 33 slices covering the entire brain, slice thickness = 3.5 mm, voxel size = 3.5 × 3.5 × 4.2 mm³.

fMRI data were preprocessed in SPM12 (Wellcome Department of Imaging Neuroscience, University College London; <http://www.fil.ion.ucl.ac.uk/spm>). Images were slice-time corrected, motion-corrected, and normalized to the Montreal Neurological Institute (MNI) space for each individual with a spatial resolution of 3 × 3 × 3 mm³. Images were then smoothed using an isotropic 6 mm Gaussian kernel and high-pass filtered at a cut-off of 128 s.

Statistical analysis

Univariate analysis

General linear model (GLM) 1. We ran a parametric analysis that included all the trials with parametric modulation, which indicates the strength of the social framing effect on a trial-by-trial basis. Specifically, we collapsed the decision phase of all trials (including both frame conditions) into one regressor. The presentations of victim information and feedback phase were also included to control related blood-oxygen-level-dependent (BOLD) responses. Parametric modulations of “to what extent the participant’s behavior was consistent with the framing effect” (i.e., the weight of framing effect for every single trial) and reaction time were added to the first regressor (i.e., the decision phase). The weight of each trial was set up as follows: for the Harm frame condition, since “not to harm” was the option that was consistent to the framing effect and “to harm” was the option that ran counter to it, we therefore assigned 4 to the 90% probability toward the “not to harm” side and −4 to the opposite case. Choosing 50% was always assigned as 0. For the Help frame condition, the 90%

probability toward the side that was consistent with the framing effect (i.e., not to help) was assigned with 4 and the other side was assigned with -4 .

This GLM analysis can reveal the strength of framing effect on a trial-by-trial basis. As the location of the avatar at the beginning of the decision phase was randomly set across trials, parametric modulation of the number of button-presses in each trial was also added according to the corresponding regressor to exclude linear signal changes of finger movement. Six parameters generated during motion correction were entered as covariates. Low-frequency drifts in signal were removed using a high-pass filter with a cutoff at 128 s. Parameter-estimated images for individual participants were calculated across the whole brain.

In addition, we ran an additional analysis to better illustrate the neural correlates of the social framing effect for two frame conditions separately (i.e., the Harm frame and Help frame) on a group level. At the first level, we built a GLM that included all the trials under the Harm frame and the Help frame condition, respectively. Other settings of this GLM were consistent with the aforementioned GLM 1. At the second level, we set up a multiple regression analysis where the individual β maps were regressed against individual difference of the social framing effect for each frame condition.

A random-effects analysis for the entire group was performed in SPM12 on the parameter-estimated images for all participants. The significance level was set to $p < 0.001$ uncorrected at the voxel level and to an extent threshold of $p < 0.05$ with false discovery rate correction (FDR) at the cluster level.

GLM 2. GLM 2 was inspired by De Martino et al. (2006), which calculated the economic framing effect as the contrast of $[Gain\ frame_{sure} + Loss\ frame_{gamble}] > [Gain\ frame_{gamble} + Loss\ frame_{sure}]$. Since the classic approach of De Martino et al. (2006) has been widely applied in neuroimaging studies (e.g., Roiser et al., 2009; Xu et al., 2013; Li et al., 2017), using a similar contrast helps compare our results with previous findings about the non-social framing effect.

The framing effect on the neural level was calculated as the contrast of $[Harm\ frame_{not\ to\ harm} + Help\ frame_{not\ to\ help}] > [Harm\ frame_{to\ harm} + Help\ frame_{to\ help}]$; that is, the difference between trials in which participant preference was affected by the frame (prefer “not to harm” in the “Harm frame” condition or prefer “not to help” in the “Help frame” condition) and trials in which participant preference ran counter to the frame (prefer “to harm” in the “Harm frame” condition or prefer “to help” in the “Help frame” condition).

Two regressors (i.e., $[Harm\ frame_{not\ to\ harm} + Help\ frame_{not\ to\ help}]$ and $[Harm\ frame_{to\ harm} + Help\ frame_{to\ help}]$) were included accordingly. Participant preference was used to categorize the data. For instance, if they selected a possibility $> 50\%$ for “not to harm” in a Harm frame trial, then this trial would be counted into the $Harm\ frame_{not\ to\ harm}$ category. The trials in which participants showed no preference (i.e., selecting 50%) were excluded from these two regressors; these trials, as well as the presentations of victim information and feedback phase, were included to control related BOLD responses.

This binary contrast GLM 2 also provides a basis for the follow-up multivoxel pattern analysis (MVPA) and psychophysiological interaction analysis (PPI) as they both require separating the trials into two comparable categories. However, this classification method disregarded the trials in which participants showed no preference between two outcomes (i.e., selecting 50%), and it might have brought the issue of imbalanced number

of trials across experimental conditions. While the GLM 2 might not be as precise as the GLM 1 methodologically, their results showed highly similar patterns (see Results).

Region of interest (ROI) analysis

Activity intensity within ROIs associated with the social and economic framing effects. To better address the findings derived from the whole-brain analysis, we performed further analyses for some predefined ROIs. We hypothesized that the social framing effect may be induced by different levels of involvement of specific regions during decision-making, such as mentalizing. To test this hypothesis, we selected ROIs from a previous meta-analysis concerning mentalizing/theory of mind, namely, MNI coordinate $[-55, -59, 19]$ for the left TPJ and MNI coordinate $[49, -56, 19]$ for the right TPJ (rTPJ) (Schurz et al., 2014). A good way to further support our viewpoint (that the social framing effect is distinct from the economic framing effect reported in the literature) is to run ROI analysis with the ROIs closely related to the economic framing effect (i.e., bilateral amygdala: $[-30, 6, -20]$ for the left and $[22, -6, -14]$ for the right; the dorsal ACC [dACC]: $[6, 26, 42]$) (adapted from Xu et al., 2013) to compare with ROIs revealed in the social framing effect. The mean parameter estimates were calculated within the sphere (radius = 6 mm) centered in the bilateral TPJ, as well as in the bilateral amygdala and dACC for each participant in the parametric analysis of the strength of social framing effect (GLM 1). We then performed one-sample t tests on the extracted values to examine whether they were significantly different from zero. We predicted that only in the bilateral TPJ, the parametric estimates would be significantly different from zero. In contrast, for the ROIs associated with the economic framing effect (bilateral amygdala and dACC), their activations should not reach significance. If that was the case, it could be regarded as evidence that the two kinds of framing effects (social vs nonsocial/economic) recruited distinct neural correlates.

Classification within ROIs. To further explore whether the activation pattern was different between trials in which participant preference was consistent with the framing effect and trials in which participant preference ran counter to the framing effect, we performed MVPA within the predefined ROIs (i.e., MNI coordinate $[-55, -59, 19]$ for the left TPJ and MNI coordinate $[49, -56, 19]$ for the rTPJ) (see Schurz et al., 2014). This method reduced the number of voxels used for classification (and therefore the number of free parameters that can lead to overfitting; for similar approaches, see Kuhl et al., 2011; Murphy et al., 2017). The classifier was trained and tested on individual parameter estimates image for $[Harm\ frame_{not\ to\ harm} + Help\ frame_{not\ to\ help}]$ and $[Harm\ frame_{to\ harm} + Help\ frame_{to\ help}]$ obtained in GLM 2.

We first defined a sphere (radius = 6 mm) around a given voxel within the ROIs (Haynes et al., 2007; Bode and Haynes, 2009). For each voxel within the ROI, we used a linear support vector machine (LIBSVM; with default hyperparameters). We adapted a leave-pair-out cross-validation procedure; thus, in the model, we have 54 ($27 + 27$) samples (participants) and 33 features (voxels). Although the model might suffer from limited sample size ($n = 54$), similar sample sizes have been commonly reported in previous studies (e.g., Shirer et al., 2012; Wang et al., 2016), especially in clinical research (e.g., Welsh et al., 2013; Dyrba et al., 2015).

The significance level (i.e., p value) of decoding accuracies was estimated by permutation tests (Tusche et al., 2016). We performed classifications with shuffled labels for 1000 times to

obtain a null distribution. The formula to compute p values is as follows: $p = (\text{ranking} + 1)/(\text{number of permutation})$.

Functional connectivity analysis (PPI). Given that the bilateral TPJ was found to play an important role in the social framing effect (see Results), we aimed to further explore whether and how the bilateral TPJ interacts with brain regions involved in decision-making. For this purpose, we performed PPI analysis (Friston et al., 1997) using the bilateral TPJ identified in our whole-brain univariate analysis (GLM 2) as seed regions, namely, MNI $[-45, -63, 27]$ for the left TPJ and MNI $[60, -63, 12]$ for the rTPJ separately. For each participant, the mean time series for each of the two seed regions were extracted and adjusted using the F -contrast for the regressors. We first set the coordinate of the peak voxel in each seed region as a landmark and searched for individual peak voxels that survived the $p < 0.05$ threshold around the landmark within a 6 mm radius sphere in each ROI. Each GLM included the following PPI regressors: (1) the main effect of seed-region activity, (2) the main effect of the contrast of $[\text{Harm frame}_{\text{not to harm}} + \text{Help frame}_{\text{not to help}}] > [\text{Harm frame}_{\text{to harm}} + \text{Help frame}_{\text{to help}}]$, and (3) their interaction. These regressors corresponded to PPI.Y, PPI.P, and PPI.ppi in the design matrix. The six head-motion parameters were also entered into the GLM as covariates to regress out any head-motion artifacts. Low-frequency drifts in signal were removed using a high-pass filter with a cutoff at 128 s. After model estimation, the PPI.ppi regressor for the contrast of $[\text{Harm frame}_{\text{to harm}} + \text{Help frame}_{\text{to help}}]$ and $[\text{Harm frame}_{\text{not to harm}} + \text{Help frame}_{\text{not to help}}]$ was entered into a random-effects analysis for the entire group. The significance level was set to $p < 0.001$ uncorrected at the voxel level and to an extent threshold of $p < 0.05$ with FDR correction at the cluster level.

Results

Behavioral results

As we stated previously, making more costly helping decisions in the Harm frame compared with the Help frame condition should result in a significant social framing effect. We used the weight value defined in GLM 1 (-4 to 4) to quantify the framing effect strength. Here, the direction of this weight value was consistent with the preference to “not harming others” in the Harm frame condition, whereas it ran counter to the preference to “helping others” in the Help frame condition. Therefore, if there was no significant difference in decision preference between the two frame conditions, the sum of weight from all trials (including both conditions) should not be different from zero. A two-sided one-sample t test was performed to examine whether the framing effect strength was significantly different from zero on the group level, and the results confirmed our hypothesis (framing effect size: 0.53 ± 0.12 [mean \pm SE], $t_{(26)} = 4.36$, $p < 0.001$) (Fig. 2).

Another two-sided paired-sample t test was performed to compare the reaction time between the two conditions, which showed no significant effect (Harm frame: 2629.58 ± 55.13 ms; Help frame: 2648.81 ± 61.48 ms [mean \pm SE], $t_{(26)} = -0.54$, $p = 0.59$).

Univariate analysis

GLM 1

The univariate analysis was aimed to reveal how brain activities were covaried with the behavioral strength of framing effect on a trial-by-trial basis within all experimental trials. The parametric modulation analysis revealed significant activations in the right calcarine gyrus (peak MNI $[18, -84, 3]$, cluster size = 349,

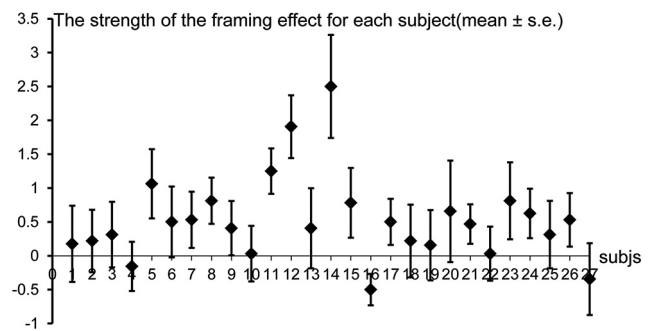


Figure 2. Behavioral results. The strength of the social framing effect for each participant (mean \pm SE).

$t_{(26)} = 8.09$), right cerebellum (peak MNI $[6, -57, -51]$, cluster size = 83, $t_{(26)} = 5.75$), left cuneus (peak MNI $[-9, -63, 27]$, cluster size = 80, $t_{(26)} = 5.60$), rTPJ (peak MNI $[51, -54, 27]$, cluster size = 53, $t_{(26)} = 5.54$), left postcentral gyrus (peak MNI $[-39, -30, 63]$, cluster size = 154, $t_{(26)} = 5.21$), and left TPJ (peak MNI $[-42, -48, 24]$, cluster size = 74, $t_{(26)} = 5.16$) (Fig. 3A; Table 1).

Both the multiple regression analysis under the Harm frame condition and that under the Help frame condition revealed significant activations in the rTPJ (Harm frame: peak MNI $[48, -75, 39]$, cluster size = 24, $t_{(26)} = 4.92$); Help frame: peak MNI $[51, -42, 45]$, cluster size = 22, $t_{(26)} = 4.24$) (Fig. 3A). The significance level for this multiple regression analysis was set at $p < 0.001$ uncorrected at the voxel level, not survived FDR correction $p < 0.05$ at the cluster level.

GLM 2

To facilitate further MVPA and PPI analysis, as well as to draw a direct comparison with previous findings about the economic framing effect (e.g., De Martino et al., 2006; Xu et al., 2013), we also ran the contrast of $[\text{Harm frame}_{\text{not to harm}} + \text{Help frame}_{\text{not to help}}] - [\text{Harm frame}_{\text{to harm}} + \text{Help frame}_{\text{to help}}]$. We found that the contrast revealed significant activation in the rTPJ (peak MNI $[60, -63, 12]$, cluster size = 62, $t_{(26)} = 4.96$), right calcarine gyrus (peak MNI $[18, -81, 3]$, cluster size = 187, $t_{(26)} = 6.40$), and left precentral gyrus extending to the postcentral gyrus (peak MNI $[-42, -24, 66]$, cluster size = 82, $t_{(26)} = 5.24$), and left TPJ (peak MNI $[-45, -63, 27]$, cluster size = 41, $t_{(26)} = 4.54$) (Fig. 3B; Table 2).

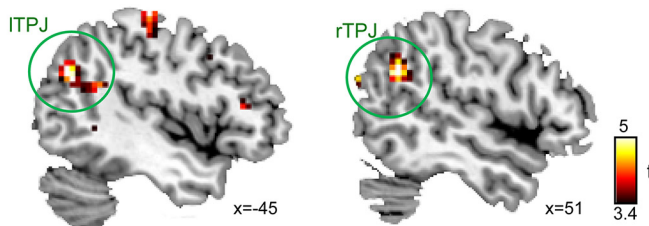
ROI-based analysis

Activity intensity within ROIs revealed in social and economic framing effects

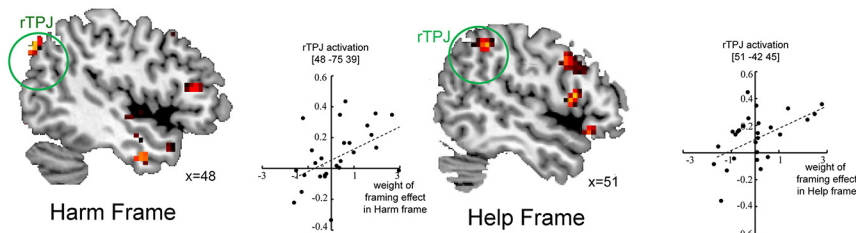
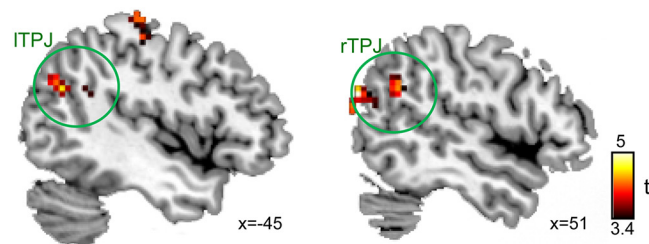
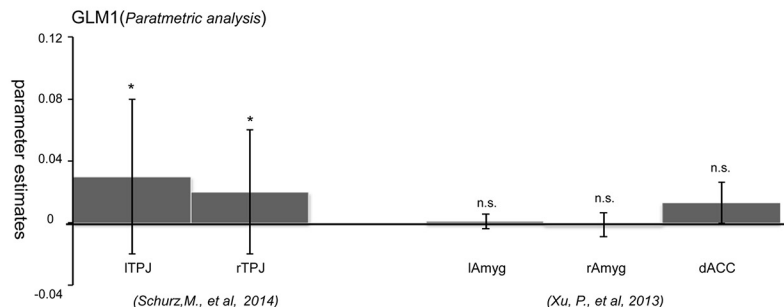
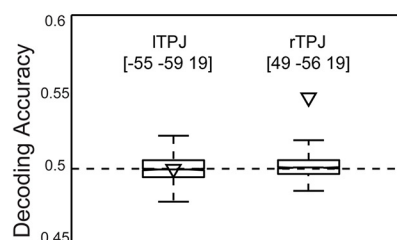
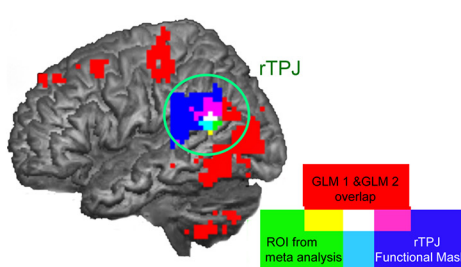
Regarding the brain regions revealed in the social framing effect, two-sided one-sample t tests revealed that the activation revealed in GLM 1 was significantly larger than zero (left TPJ: 0.03 ± 0.05 , $t_{(26)} = 2.68$, $p = 0.01$; rTPJ: 0.02 ± 0.04 , $t_{(26)} = 2.77$, $p = 0.01$). For the regions revealed in the economic framing effect, the bilateral amygdala and dACC showed no differences with zero (left amygdala: $t_{(26)} = 0.30$, $p = 0.77$; right amygdala: $t_{(26)} = -0.14$, $p = 0.89$; dACC: $t_{(26)} = 1.04$, $p = 0.31$) (Fig. 3C).

Classification within ROIs

We found that the rTPJ could significantly classify between $[\text{Harm frame}_{\text{not to harm}} + \text{Help frame}_{\text{not to help}}]$ trials and $[\text{Harm frame}_{\text{to harm}} + \text{Help frame}_{\text{to help}}]$ trials in the $1000 \times$ permutation (classification accuracy = 54.47%, $p = 0.001$), but the left TPJ could not (classification accuracy = 49.93%, $p > 0.05$) (Fig. 3D). To demonstrate that the activations of rTPJ revealed in GLM 1,

A GLM1 parametric analysis

Group-level difference of framing effect for Harm and Help frames

**B** GLM2 [*Harm frame* not to harm + *Help frame* not to help] minus [*Harm frame* to harm + *Help frame* to help]**C** ROI Analysis**D** Classification within ROIs**E** rTPJ revealed in GLM1&2 and ROI analysis

GLM 2, and the rTPJ ROI selected in meta-analysis are converging within the rTPJ region, we overlapped these images together with a functional rTPJ mask (see also Bzdok et al., 2013) (Fig. 3E).

PPI analysis

We performed PPI analyses to test whether the functional connectivity between the bilateral TPJ and other brain regions was modulated by frame manipulation. For the seed region of the left TPJ, we did not observe any difference in functional connectivity between the [*Harm frame* not to harm + *Help frame* not to help] trials and the [*Harm frame* to harm + *Help frame* to help] trials. However, for the seed region of the rTPJ, we found significant difference in functional connectivity of the medial prefrontal cortex (mPFC) for [*Harm frame* to harm + *Help frame* to help] trials > [*Harm frame* not to harm + *Help frame* not to help] trials (peak MNI [15, 45, 18], cluster size = 110, $t_{(26)} = 5.16$) (Fig. 4A).

To explore whether brain functional connectivity was associated with the observed social framing effect, we calculated the correlation between the functional connectivity strength of rTPJ to mPFC and the behavioral social framing effect. Pearson's correlation analysis revealed that the functional connectivity between the rTPJ and mPFC predicted the social framing effect size ($r = 0.39$, $p = 0.04$) (Fig. 4B).

Experiment 2 (tDCS)**Materials and methods****Participants**

Sixty-two right-handed participants were recruited from local universities for participation in the tDCS experiment. None of these persons had participated in Experiment 1. Two participants were excluded due to their absence on the second experimental day, leaving 60 participants in the final sample.

All participants were screened for histories of neurologic disorders, brain injury, or developmental disabilities. All had normal or corrected-to-normal vision. The study was conducted according to the ethical guidelines and principles of the Declaration of Helsinki and was approved by the Medical Ethics Committee of Shenzhen University

Figure 3. Results of brain activation. **A**, Results of GLM 1 (top) and the results of multiple regression analysis separately for two frame conditions (bottom), showing that the rTPJ activation was related to the social framing effect. **B**, Results of the traditional binary analysis (GLM 2) further support the findings from GLM 1 that the rTPJ underlies the social framing effect. **C**, Results of ROI analysis (mean \pm SE). **D**, MVPA (i.e., classification activation differences based on the contrast of GLM 2; the whiskers in the box-plot show the maximum and minimum values). **E**, 3D overlap of the rTPJ revealed in **A–C** and the functional mask of rTPJ extracted from (Bzdok et al., 2013). All imaging maps are thresholded at $p < 0.001$ uncorrected. * $p < 0.05$, n.s.: not significant.

Table 1. Whole-brain activations based on GLM 1

Brain region	BA	Coordinates			Volume	T
		(x)	(y)	(z)		
Right calcarine	18	18	−84	3	3.49	8.09
	17	9	−81	15		6.05
	18	15	−63	−3		5.61
Right cerebellum		6	−57	−51	83	5.75
		15	−51	−48		5.66
		6	−48	−54		4.61
Left cuneus	23	−9	−63	27	80	5.60
		−18	−57	24		5.22
		−15	−45	27		4.59
Right TPJ	22	51	−54	27	53	5.54
	41	45	−48	24		4.31
Left postcentral		−39	−30	63	154	5.21
		−45	−21	63		5.01
Left TPJ	41	−42	−48	24	74	5.16
	39	−45	−63	27		5.15
	21	−51	−57	21		4.90

Table 2. Whole-brain activations based on GLM 2

Brain region	BA	Coordinates			Volume	T
		(x)	(y)	(z)		
Right calcarine gyrus	18	18	−81	3	187	6.40
Right lingual gyrus	19	18	−63	−6		5.37
Left precentral/postcentral gyrus	4	−42	−24	66		5.24
	4	−39	−27	69	82	5.06
	3	−33	−30	54		4.38
Right TPJ	37	60	−63	12	62	4.96
	39	51	−75	24		4.60
	37	57	−66	9		4.59
Left TPJ	39	−45	−63	27	27	4.54

Medical School. All participants provided written informed consent after they fully understood the procedures.

Experimental design and statistical analysis

To test the causal role of the rTPJ in the social framing effect (compared with the nonsocial one), we performed two experimental sessions using tDCS. To better control individual difference, which might strongly influence the results of tDCS experiment, we used a novel way developed by Zhang et al. (2016). The sample was separated into two groups, one of which ($n = 30$, 15 women, age: 21.9 ± 1.9 years) received anodal and sham stimulation over the rTPJ in two experimental sessions on 2 different days (separated by 1 week); the time sequence of receiving anodal and sham stimulation was counterbalanced across participants. The second group ($n = 30$, 15 women, age: 20.87 ± 2.3 years) received cathodal and sham stimulation over the same region in two experimental sessions, the time sequence of which was also counterbalanced across participants. In this way, the effect of anodal stimulation can be directly compared with that of sham stimulation within the same group. Likewise, the effect of cathodal stimulation can be directly compared with that of sham stimulation within the same group. Since these comparisons were both conducted on the within-subject level, the effect of individual difference would be better controlled.

Each experimental session included two tasks (i.e., social and nonsocial framing). The social framing task was the same as that in Experiment 1, but its trial number was reduced to 25 per condition (Harm frame vs Help frame). The nonsocial framing task was adapted from De Martino et al. (2006), in each trial of which participants were asked to select between a sure option (i.e., only

retain a portion of the reward) and a gambling option; the expected values of the two options were identical within each trial, but the sure option was described as either a gain or a loss. We excluded the original control condition (“catch” trials: the expected values of two options were unbalanced) and retained 32 trials for each condition (Gain frame vs Loss frame). The trial number was modified to shorten the total amount of time of the tDCS experiments. Consequently, each task lasted for 8 min, and there was a 1.5 min rest between them. The order of the two tasks was counterbalanced across participants. After the experiment, participants received remuneration (~ 100 RMB) according to their task performance.

tDCS parameters

A constant electrical current of 2 mA was applied by a saline-soaked pair of surface sponge rubber electrodes and delivered by a battery-driven transcranial stimulator (Brainstim, EMS). It was adjusted to induce cortical excitability of the target area without any physiological damage. tDCS started 8 min before the task and was delivered during the whole course of the experiment. For sham stimulation, the electrodes were placed in the same positions as for the anodal/cathodal stimulation, but the stimulator was only turned on for the initial 30 s. This method of sham stimulation has been proven to be reliable (Gandiga et al., 2006; Nihonsugi et al., 2015).

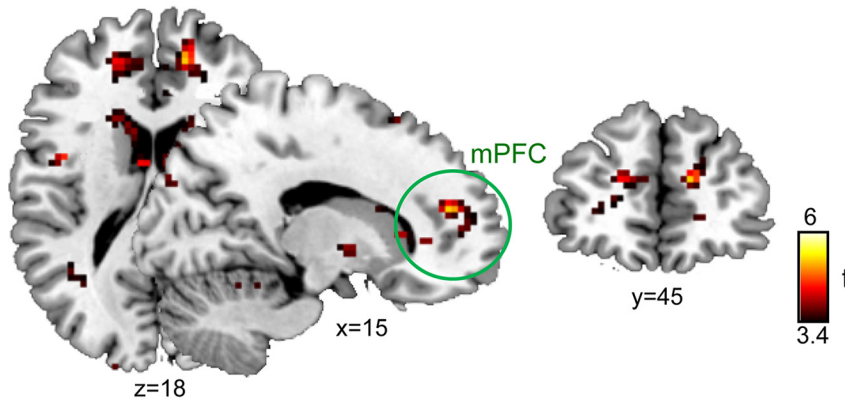
The stimulation site in our study was localized using the EEG 10/20 system. A 5×5 cm electrode was placed over the rTPJ area (at the 10-20 electrode positions between CP6 and P6) while a reference electrode was positioned above the left eyebrow. Both electrodes were positioned firmly on the scalp with a plastic headband. The electrode montage and tDCS parameters were identical to those that successfully modulated cortical excitability of the rTPJ in previous studies (Jurcak et al., 2007; Santiesteban et al., 2012; Tang et al., 2017).

Results

For the social framing task, the framing effect size was calculated in the same way as Experiment 1. Consistent with the behavioral findings in Experiment 1, one-sample t test revealed that the social framing effect was significantly larger than zero (anodal/sham group: $t_{(29)} = 4.02$, $p < 0.001$ for anodal stimulation; $t_{(29)} = 2.71$, $p = 0.011$ for sham stimulation; cathodal/sham group: $t_{(29)} = 2.05$, $p = 0.047$ for cathodal stimulation; $t_{(29)} = 2.46$, $p = 0.020$ for sham stimulation), indicating that the frame manipulation was effective. We then performed paired-sample t test between anodal/cathodal and sham stimulation for each group separately. For the social framing task, the framing effect size significantly increased under anodal stimulation compared with sham stimulation (anodal vs sham: $t_{(29)} = 2.24$, $p = 0.03$); meanwhile, the framing effect size decreased under cathodal stimulation compared with sham stimulation, which approached statistical significance (cathodal vs sham: $t_{(29)} = -1.86$, $p = 0.07$) (Fig. 5A).

For the nonsocial framing task, consistent with the original reference (De Martino et al., 2006), the framing effect size in the nonsocial framing task was calculated as $(Gain\ frame_{sure} + Loss\ frame_{gamble}) - (Gain\ frame_{gamble} + Loss\ frame_{sure})$. One-sample t test revealed that the nonsocial framing effect was significantly larger than zero (anodal/sham group: $t_{(29)} = 2.20$, $p = 0.036$ for anodal stimulation; $t_{(29)} = 4.75$, $p < 0.001$ for sham stimulation; cathodal/sham group: $t_{(29)} = 2.55$, $p = 0.016$ for cathodal stimulation; $t_{(29)} = 3.60$, $p = 0.001$ for sham stimulation), indicating that the frame manipulation was effective. However, the difference

A PPI results



B Functional connectivity predicts the framing effect

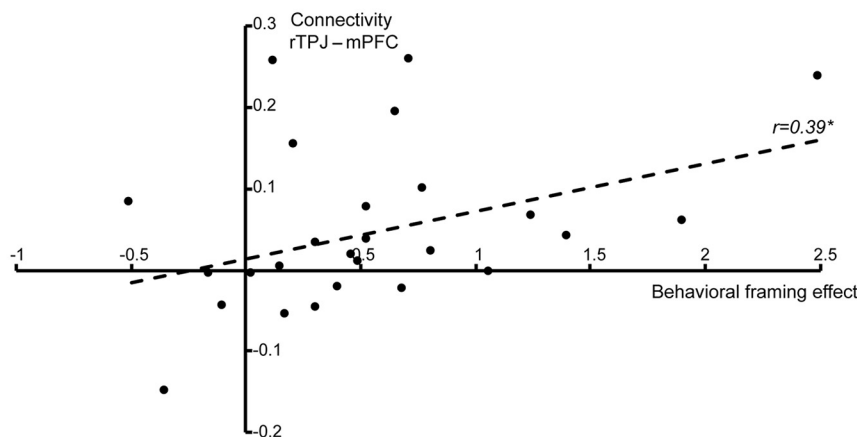


Figure 4. Results of PPI. **A**, Results of PPI. **B**, The strength of the functional connectivity between the rTPJ and mPFC predicted the behavioral social framing effect size. * $p < 0.05$.

in framing effect size between anodal/cathodal stimulation and sham stimulation was nonsignificant (anodal vs sham: $t_{(29)} = 1.06$, $p = 0.30$; cathodal vs sham: $t_{(29)} = 0.35$, $p = 0.73$) (Fig. 5B), indicating that the nonsocial framing effect was insensitive to tDCS stimulation on the rTPJ (Fig. 5B).

Regarding the reaction time results, we detected no significant effect in either task.

Discussion

In this study, we explored the neural circuit of the social framing effect, which might involve important brain regions related to social cognition. A novel task has been developed for this purpose, in which participants may maximize reward at the expense of exposing their victim to the risk of receiving a painful shock. In the Harm frame condition, the shock is described as if it is the consequence of the participants' own action; in the Help frame condition, it is implied that the shock delivered is part of the task, but participants could help the victim to avoid it. Since avoiding harming others is a strong, universal motive among human beings, we predicted that, compared with the Help frame condition, participants would be more likely to select the costly helping option in the Harm frame condition.

Consistent with this hypothesis, the behavioral results in Experiment 1 revealed that participant preference to the costly helping option significantly increased in the Harm frame

condition than in the Help frame condition, which was analogous to the classic social framing effect. On the brain-imaging level, the (bilateral) TPJ, (right) calcarine gyrus, (left) cuneus, (right) cerebellum, and (left) postcentral gyrus were more activated as a function of the social framing effect size. We also examined neural activations that covaried with the framing effect on a group level. The results showed that, in both frame conditions, the BOLD activation of the rTPJ increased when participants showed a stronger framing effect. Follow-up permutation test revealed that the rTPJ and its functional connectivity with the mPFC showed a more reliable relationship with the social framing effect than its left counterpart. Given that brain-imaging techniques are constrained in making causal inference (Ramsey et al., 2010), we conducted a separate tDCS experiment to modulate rTPJ activity. Results of this Experiment 2 show that, compared with sham stimulation, the social framing effect was stronger under anodal stimulation but became weaker under cathodal stimulation (marginally significant). In contrast, the nonsocial framing effect in the gambling task was insensitive to tDCS manipulation. Overall, we confirmed the importance of the TPJ to the social framing effect.

Previous studies have shown that a key function of the TPJ is to differentiate one's own from the other's perspective (Frith and Frith, 2010; Speitel et al., 2019). TPJ activation has been frequently reported in social decision-making tasks, especially when those tasks engage other-regarding considerations (Morishima et al., 2012; Tusche et al., 2016; Park et al., 2017). In addition, the current results revealed that the rTPJ weighs more heavily on the social framing effect than does its left counterpart. This hemispheric asymmetry has also been supported by the literature, as the rTPJ has been more strongly related to social processes (Krall et al., 2015). In our opinion, the decision "to harm" is more morally reprehensible than "not to help," which may trigger a stronger conflict between moral values and material concerns. Previous studies have indicated that the rTPJ may be responsible for representing this conflict in decision-making (Berns et al., 2012; Morishima et al., 2012). Compared with the Help frame, the Harm frame in our task emphasized the "harmful" consequences to others and therefore triggered a tendency to avoid moral conflicts. As a result, the participants were more likely to avoid shocking others at the expense of their own gain.

An alternative interpretation is associated with the idea that the rTPJ is specialized for reasoning regarding the affective and cognitive mental states of others (Saxe, 2006; Zhu et al., 2019). It is possible that, under the Harm frame condition,

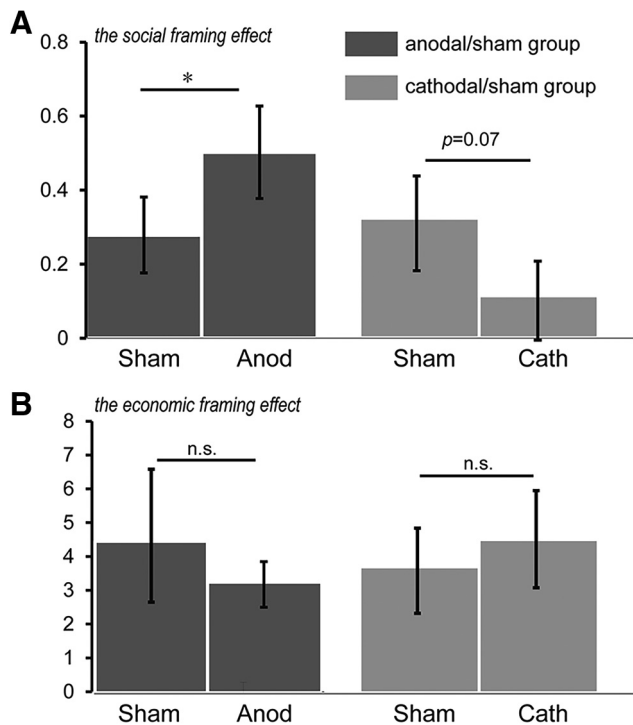


Figure 5. Results of tDCS experiments: social framing effect (mean \pm SE). **A**, Left to right, The strength of social framing effect in the anodal/sham group and in the cathodal/sham group. **B**, Left to right, The strength of economic framing effect in the anodal/sham group and in the cathodal/sham group. * $p < 0.05$. n.s.: not significant.

the participants experienced a stronger inferred feeling state of their victim associated with a shock, which in turn strengthened the feeling of guilt. Consequently, behavioral preference varied as a function of the sensitivity to physical pain of the victim across conditions. While follow-up research is needed to determine which interpretation is more appropriate, both of them are in line with the finding that activating the rTPJ decreases antisocial behavior (e.g., deception) (see Tang et al., 2017).

Regarding the social nature of the current paradigm, here we interpret the major findings from the perspective of social functions. However, it is undeniable that the rTPJ has also been associated with nonsocial functions, such as attentional reorienting and reallocation (for review, see Krall et al., 2015). Accordingly, one might suggest that rTPJ activations in the current fMRI data actually indicate the importance of attentional function to our social framing task. That is, the social frames influenced behavioral preference by modulating attentional allocation. But this hypothesis might encounter difficulties to explain why tDCS stimulation on the rTPJ did not affect individual decisions in the nonsocial framing (Gain/Loss) task as well, since previous studies have shown that attention allocation also plays an important role in economic risk-taking (e.g., Yechiam and Hochman, 2014; Zhang and Gu, 2018).

Further, the results of the PPI analysis highlighted the connection between the rTPJ and mPFC. As one of the cortical hubs, the mPFC is a key region for information integration and valuation (Knutson et al., 2007; McNamee et al., 2013). Using the gambling task, one of our previous studies found that the mPFC was positively coupled with the bilateral

amygdala under the nonsocial framing effect context, indicating the influence of emotional information from the amygdala to the mPFC (Xu et al., 2013). The current study also showed that the strength of the functional connectivity between the rTPJ and mPFC predicted the social framing effect, which indicates the importance of the connection between the mPFC and TPJ (instead of the amygdala) under social contexts. It is worth noting that the TPJ and mPFC are involved in the same mentalizing network (Bault et al., 2011; Mars et al., 2012). The significance of the TPJ-mPFC connection in the social framing task may indicate the mechanism of how social information transfers between the TPJ and the mPFC to bias decision preference when people receive the same offers that are framed differently.

Aside from the way of framing, the two options in different frame conditions (i.e., “not to harm” vs “to help”) also differed in the level of agency recruitment. That is, the option “not to harm” was worded as an inaction, whereas the option “to help” was worded as an action. Regarding that, the current results might have been affected by the “omission bias,” which refers to a tendency to avoid regret caused by actions compared with regret caused by inactions (Baron and Ritov, 2004). We admit this possibility but suggest that the experimental effects associated with the omission bias were limited for two reasons. First, the difference between the two options in the level of agency recruitment was not emphasized in our task instruction. Second, seeing that actively moving the avatar to modulate outcome probabilities could be understood as an “action” while leaving the avatar at the starting point could be understood as an “inaction,” the influence of the omission bias might actually be balanced across conditions. Nonetheless, it would be necessary for future studies to use alternative paradigms to better control the omission bias.

To sum up, using fMRI and tDCS, respectively, the two experiments in the current study consistently showed that the social framing effect was strongly associated with the rTPJ. Stronger rTPJ activations might reflect a higher level of moral conflict and/or mentalizing with others when the social responsibility for participants regarding the shock was emphasized by the option frame (i.e., the Harm frame). Consequently, the participants were more prone to “harm aversion” and selected costly helping. These findings may shed light on the issue of enhancing prosocial-helping behavior by manipulating verbal information (i.e., nudges; see Coventry et al., 2016).

Finally, some limitations should be noted. Regarding the technical issues, one drawback of the MVPA in the current study is that the sample size was relatively small for machine learning algorithm, although some previous studies used similar sample sizes (e.g., Shirer et al., 2012; Wang et al., 2016). Moreover, the results of the second-level multiple regression analysis did not survive FDR correction and the two rTPJ clusters revealed in the Harm and the Help frame conditions did not overlap, also possibly due to the relatively limited sample size. We should consider these results with caution and further verification with a larger sample size would be needed. In addition, the paradigm used in this study was different from those in previous studies on the social framing effect (e.g., the Community Game vs Wall Street Game frame in economic cooperation games) (see Liberman et al., 2004). We did not choose those paradigms partly because of the heterogeneous findings in the literature (e.g., Rege and Telle, 2004; Cubitt et al., 2011; Dreber et al., 2011; Ellingsen et al.,

2012). Also, we were interested in investigating the influence of social frames on different decision problems beyond economic cooperation. We believe that the current findings have general implications in the brain mechanisms of the social framing effect, but follow-up examinations would surely be necessary.

References

- Aupperle RL, Sullivan S, Melrose AJ, Paulus MP, Stein MB (2011) A reverse translational approach to quantify approach-avoidance conflict in humans. *Behav Brain Res* 225:455–463.
- Aupperle RL, Melrose AJ, Francisco A, Paulus MP, Stein MB (2015) Neural substrates of approach-avoidance conflict decision-making. *Hum Brain Mapp* 36:449–462.
- Baron J, Ritov I (2004) Omission bias, individual differences, and normality. *Org Behav Hum Decis Process* 94:74–85.
- Bault N, Joffily M, Rustichini A, Coricelli G (2011) Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proc Natl Acad Sci USA* 108:16044–16049.
- Berns GS, Bell E, Capra CM, Prietula MJ, Moore S, Anderson B, Ginges J, Atran S (2012) The price of your soul: neural evidence for the non-utilitarian representation of sacred values. *Philos Trans R Soc Lond B Biol Sci* 367:754–762.
- Bode S, Haynes JD (2009) Decoding sequential stages of task preparation in the human brain. *Neuroimage* 45:606–613.
- Bzdok D, Langner R, Schilbach L, Jakobs O, Roski C, Caspers S, Laird AR, Fox PT, Zilles K, Eickhoff SB (2013) Characterization of the temporoparietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage* 81:381–392.
- Coventry LM, Jeske D, Blythe JM, Turland J, Briggs P (2016) Personality and social framing in privacy decision-making: a study on cookie acceptance. *Front Psychol* 7:1341.
- Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, Dolan RJ (2014) Harm to others outweighs harm to self in moral decision making. *Proc Natl Acad Sci USA* 111:17320–17325.
- Cubitt RP, Drouvelis M, Gächter S (2011) Framing and free riding: emotional responses and punishment in social dilemma games. *Exp Econ* 14:254–272.
- De Martino B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. *Science* 313:684–687.
- Dreber A, Ellingsen T, Johannesson M, Rand DG (2011) Do people care about context? Framing effects in dictator games. In: *SSE/EFI Working Paper Series in Economics and Finance* (Stockholm School of Economics TERIE, ed). Stockholm.
- Dyrba M, Grothe M, Kirste T, Teipel SJ (2015) Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Hum Brain Mapp* 36:2118–2131.
- Ellingsen T, Johannesson M, Mollerstrom J, Munkhammar S (2012) Social framing effects: preferences or beliefs? *Game Econ Behav* 76:117–130.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
- Frith U, Frith C (2010) The social brain: allowing humans to boldly go where no other species has been. *Philos Trans R Soc Lond B Biol Sci* 365:165–176.
- Gandiga PC, Hummel FC, Cohen LG (2006) Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin Neurophysiol* 117:845–850.
- Guitart-Masip M, Talmi D, Dolan R (2010) Conditioned associations and economic decision biases. *Neuroimage* 53:206–214.
- Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. *Curr Biol* 17:323–328.
- Jurcak V, Tsuzuki D, Dan I (2007) 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* 34:1600–1611.
- Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53:147–156.
- Krall SC, Rottschy C, Oberwelland E, Bzdok D, Fox PT, Eickhoff SB, Fink GR, Konrad K (2015) The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Struct Funct* 220:587–604.
- Kühberger A (1998) The influence of framing on risky decisions: a meta-analysis. *Org Behav Hum Decis Process* 75:23–55.
- Kuhl BA, Rissman J, Chun MM, Wagner AD (2011) Fidelity of neural reactivation reveals competition between memories. *Proc Natl Acad Sci USA* 108:5903–5908.
- Levin IP, Schneider SL, Gaeth GJ (1998) All frames are not created equal: a typology and critical analysis of framing effects. *Org Behav Hum Decis Process* 76:149–188.
- Li R, Smith DV, Clithero JA, Venkatraman V, Carter RM, Huettel SA (2017) Reason's enemy is not emotion: engagement of cognitive control networks explains biases in gain/loss framing. *J Neurosci* 37:3588–3598.
- Liberman V, Samuels SM, Ross L (2004) The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Pers Soc Psychol Bull* 30:1175–1185.
- Mars RB, Sallet J, Schuffelgen U, Jbabdi S, Toni I, Rushworth MF (2012) Connectivity-based subdivisions of the human right “temporoparietal junction area”: evidence for different areas participating in different cortical networks. *Cereb Cortex* 22:1894–1903.
- McNamee D, Rangel A, O'Doherty JP (2013) Category-dependent and category-independent goal-value codes in human ventromPFC. *Nat Neurosci* 16:479–485.
- Morishima Y, Schunk D, Bruhin A, Ruff CC, Fehr E (2012) Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75:73–79.
- Murphy C, Rueschemeyer SA, Watson D, Karapanagiotidis T, Smallwood J, Jefferies E (2017) Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. *Neuroimage* 147:19–31.
- Nihonsugi T, Ihara A, Haruno M (2015) Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *J Neurosci* 35:3412–3419.
- Park SQ, Kahnt T, Dogan A, Strang S, Fehr E, Tobler PN (2017) A neural link between generosity and happiness. *Nat Commun* 8:15964.
- Parkinson C, Wheatley T (2015) The repurposed social brain. *Trends Cogn Sci* 19:133–141.
- Rabin M (1998) Psychology and economics. *J Econ Lit* 51:528–546.
- Ramsey JD, Hanson SJ, Hanson C, Halchenko YO, Poldrack RA, Glymour C (2010) Six problems for causal inference from fMRI. *Neuroimage* 49:1545–1558.
- Rege M, Telle K (2004) The impact of social approval and framing on cooperation in public good situations. *J Public Econ* 88:1625–1644.
- Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. *Annu Rev Psychol* 62:23–48.
- Roiser JP, de Martino B, Tan GC, Kumaran D, Seymour B, Wood NW, Dolan RJ (2009) A genetically mediated bias in decision making driven by failure of amygdala control. *J Neurosci* 29:5985–5991.
- Sanfey AG (2007) Social decision-making: insights from game theory and neuroscience. *Science* 318:598–602.
- Santesteban I, Banissy MJ, Catmur C, Bird G (2012) Enhancing social ability by stimulating right temporoparietal junction. *Curr Biol* 22:2274–2277.
- Saxe R (2006) Uniquely human social cognition. *Curr Opin Neurobiol* 16:235–239.
- Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 42:9–34.
- Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2012) Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* 22:158–165.
- Speitel C, Traut-Mattausch E, Jonas E (2019) Functions of the right DLPFC and right TPJ in proposers and responders in the ultimatum game. *Soc Cogn Affect Neurosci* 14:263–270.
- Tang H, Ye P, Wang S, Zhu R, Su S, Tong L, Liu C (2017) Stimulating the right temporoparietal junction with tDCS decreases deception in moral hypocrisy and unfairness. *Front Psychol* 8:2033.
- Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, Trope Y, Schiller D (2015) A map for social navigation in the human brain. *Neuron* 87:231–243.

- Tusche A, Bockler A, Kanske P, Trautwein FM, Singer T (2016) Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *J Neurosci* 36:4719–4732.
- Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211:453–458.
- Van Dijk E, Wilke H (2000) Decision-induced focusing in social dilemmas: give-some, keep-some, take-some, and leave-some dilemmas. *J Pers Soc Psychol* 78:92–104.
- Wang X, Fang Y, Cui Z, Xu Y, He Y, Guo Q, Bi Y (2016) Representing object categories by connections: evidence from a multivariate connectivity pattern classification approach. *Hum Brain Mapp* 37:3685–3697.
- Wang XT, Rao LL, Zheng H (2017) Neural substrates of framing effects in social contexts: a meta-analytical approach. *Soc Neurosci* 12:268–279.
- Welsh RC, Jelsone-Swain LM, Foerster BR (2013) The utility of independent component analysis and machine learning in the identification of the amyotrophic lateral sclerosis diseased brain. *Front Hum Neurosci* 7:251.
- Xu P, Gu R, Broster LS, Wu R, Van Dam NT, Jiang Y, Fan J, Luo YJ (2013) Neural basis of emotional decision making in trait anxiety. *J Neurosci* 33:18641–18653.
- Yechiam E, Hochman G (2014) Loss attention in a dual-task setting. *Psychol Sci* 25:494–502.
- Zhang D, Gu R (2018) Behavioral preference in sequential decision-making and its association with anxiety. *Hum Brain Mapp* 39:2482–2499.
- Zhang Y, Yu H, Yin Y, Zhou X (2016) Intention modulates the effect of punishment threat in norm enforcement via the lateral orbitofrontal cortex. *J Neurosci* 36:9217–9226.
- 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.