Brain Imaging and Behavior (2019) 13:685–698
https://doi.org/10.1007/s11682-018-9893-1

ORIGINAL RESEARCH

The neural signatures of egocentric bias in normative decision-making

Chunliang Feng[1,2] · Xue Feng[3] · Li Wang[4] · Lili Wang[5] · Ruolei Gu[6,7] · Aiping Ni[6,7] ·
Gopikrishna Deshpande[8,9,10] · Zhihao Li[1] · Yue-Jia Luo[1,11]

Abstract

Bargaining parties often disagree on what fair is, due to the reason that people are prone to believe that what favors oneself is fair, i.e., an egocentric bias. In this study, we investigated the neural signatures underlying egocentric bias in fairness decision-making, conjoining an adapted ultimatum game (UG) with event-related fMRI and functional connectivity. Participants earned monetary rewards with a partner in a production stage, wherein their contributions to the earnings were manipulated. Afterwards, the joint earnings were randomly divided, and the distribution was presented simultaneously with contribution information to participants, who accepted/rejected distributions of earnings as the same manner in standard UG. We identified an egocentric bias in fairness decisions, such that participants frequently rejected self-contributed disadvantageous outcomes, but much less so in response to other-contributed advantageous outcomes, although both involved mismatch between contribution and payoff. This bias was underpinned by regions involved in representing fairness norms, including the anterior insula and dorsal anterior cingulate cortex (dACC). Furthermore, the thalamus activity was predictive of the bias, such that the level of egocentric bias decreased as a function of the activation level of the thalamus. Finally, our functional-connectivity findings indicated that the thalamus worked together with insula and dACC to modulate behavioral egocentric bias in fairness-related decisions. Our findings uncover the neural basis underlying the modulation of egocentric bias in normative decision-making, and highlight the role of neural circuits associated with norm enforcement in this phenomenon.

Keywords Fairness · Egocentric bias · Self-interest · Ultimatum game · fMRI · Psychophysiological interactions

---

Chunliang Feng and Li Wang contributed equally to the current work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11682-018-9893-1) contains supplementary material, which is available to authorized users.

✉ Zhihao Li
zli8@emory.edu

✉ Yue-Jia Luo
luoyj@szu.edu.cn

[1] Shenzhen Key Laboratory of Affective and Social Cognitive Science, Shenzhen University, Shenzhen, China

[2] College of Information Science and Technology, Beijing Normal University, Beijing, China

[3] Key Laboratory of Modern Teaching Technology of Ministry of Education, Shaanxi Normal University, Xi'an, China

[4] Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing, China

[5] School of Educational Science, Huaiyin Normal University, Huai'an, China

[6] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[7] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[8] Department of Electrical and Computer Engineering, Auburn University MRI Research Center, Auburn University, Auburn, AL, USA

[9] Department of Psychology, Auburn University, Auburn, AL, USA

[10] Alabama Advanced Imaging Consortium, Auburn University and University of Alabama Birmingham, Auburn, AL, USA

[11] Center for Emotion and Brain, Shenzhen Institute of Neuroscience, Shenzhen, China

Introduction

Fairness constitutes one of the fundamental social norms across human societies. People perceive fair distributions as rewards and costly punish unfair behaviors as norm violations (Henrich et al. 2006; Buckholtz and Marois 2012; Feng et al. 2015). However, it is practically difficult for bargaining parties to agree on what fair is, since people are prone to consider fairness in an egocentric manner. That is, people tend to bias the fairness judgment in favor of one's own interest (Kriss et al. 2011). For instance, people often show an egocentric bias in interpreting contribution during fairness-related decision-making (Rodriguez-Lara and Moreno-Garrido 2012; Feng et al. 2013). Contribution is deemed as a more important factor during resource distribution when people contribute more to the earnings compared to the condition when they contribute less than others (Rutström and Williams 2000; Bediou and Scherer 2014). Furthermore, the egocentric bias in interpreting contribution has been identified when people respond to deviations from a proportional distribution. Specially, when incurring disadvantageous inequality in the context of a larger contribution (i.e., contributions outweigh payoffs), people react harshly and often reject such distributions. By contrast, people usually ignore the deviations when receiving advantageous inequality in the context of a smaller contribution (i.e., payoffs outweigh contributions) (Feng et al. 2013; Bediou and Scherer 2014). The egocentric bias pervades bargaining situations with substantial impact on fairness-relevant judgments and decision-making (e.g., Bediou and Scherer 2014; Loewenstein et al. 1993; Otto and Bolle 2015), however, the neurocognitive mechanisms underlying this phenomenon remain understudied.

Previous brain imaging studies have implicated multiple neural systems in human fairness decision-making (Sanfey et al. 2003; Feng et al. 2015). On one hand, the anterior insula (AI) and dorsal anterior cingulate cortex (dACC) may constitute a reflexive system to detect unfairness as norm violations (Chang and Sanfey 2013; Xiang et al. 2013; Chang and Smith 2015). These regions connect extensively with subcortical structures (e.g., thalamus) that are involved in generating aversive responses (e.g., punishment) to norm violations (Hu et al. 2015; Corradi-Dell'Acqua et al. 2016). On the other hand, the deliberate system consisting of the dorsolateral prefrontal cortex (dlPFC), dorsomedial PFC (dmPFC), and ventrolateral PFC (vlPFC) is thought to be responsible for reappraising and regulating the reflexive system in favor of either economic self-interest or enforcement of social norms (Buckholtz and Marois 2012; Grecucci et al. 2013; Feng et al. 2015). In particular, the right dlPFC has been implicated in facilitating punishment to norm violations by over-riding temptations of self-interest, such that its disruption diminishes rejection responses to unfair offers (Knoch et al. 2006). In contrast, the dmPFC and vlFPC are associated with suppressing aversive reactions to unfair offers due to self-interest motivations, leading to an increased acceptance of unfair offers (Tabibnia et al. 2008; Civai et al. 2012). In short, previous neuroimaging evidence indicates that fairness consideration incorporates both automatic and effortful cognitive processes. Notably, both reflexive and deliberate systems are sensitive to social context, supporting flexible fairness decisions in different situations (Güroğlu et al. 2010; Grecucci et al. 2013). It is thus plausible that egocentric bias in normative decision-making, manifested as flexible interpretations of social contexts (e.g., contribution), engages the modulation of neural responses in brain regions previously implicated in fairness judgments and decisions.

We addressed this issue in the current study by combining functional magnetic resonance imaging (fMRI) and context-dependent functional connectivity with a variant of ultimatum game (UG). Prior to the fMRI scanning, participants jointly earned monetary rewards with another same-sex partner by completing a cognitive task, during which their contributions to the earning were manipulated according to the relative performance in the task. Afterwards, participants were scanned while acting as responders and received equal or unequal distributions of the earning in the context of different contributions. On each round they needed to decide whether to accept (participant and partner got paid accordingly) or reject (neither got paid) each income distribution. Participants' self-reports of unfairness ratings were also collected for assessing the egocentric bias in fairness judgments.

We hypothesized that participants would respond differently to contributions in disadvantageous and advantageous conditions. On the one hand, participants would reject disadvantageous inequality much more frequently in the self-contribution condition than the other-contribution condition, presumably due to the reason that disadvantageous outcomes do not match their efforts in the self-contribution condition. On other hand, participants might not exhibit higher rejection rates to advantageous outcomes in the other-contribution condition (than the self-contribution condition), although these outcomes do not match their efforts either. Such differential considerations of contribution reflect egocentric biases, since they depend on the status of self-interest. This phenomenon is common in realistic life, such that people often complain about their efforts outweighing pay-offs, but rarely complain about their pay-offs outweighing efforts. At the neural level, the current work aimed to assess two potential neurocognitive mechanisms of the egocentric bias: (i) the bias is associated with attenuated response in brain network recruited by fairness encoding (e.g., AI, dACC) and weaker functional connectivity of this network with subcortical regions (e.g., thalamus) implicated in generating aversive responses to norm violations, reflecting a distorted perception of fairness norms (Thompson and Loewenstein 1992; Loewenstein et al. 1993), and/or (ii) the egocentric bias is associated with stronger neural activity of the dmPFC and vlPFC mediating self-interested decisions, reflecting the involvement of self-interest motivations (Blake et al. 2014; Otto and Bolle 2015).

Materials and methods

Subjects

Twenty-three students (12 females) (mean age ± s.d.: 22.78 ± 1.95) participated in the study for monetary compensation. The sample size was based on previous fMRI studies using a similar paradigm (Civai et al. 2012; Feng et al. 2016b). All participants were right-handed (based on self-reports), had normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders. Written informed consents were obtained from all participants. The study was conducted according to the ethical guidelines and principles of the Declaration of Helsinki and was approved by the Institutional Review Board at Beijing Normal University (BNU), Beijing, China.

Experimental procedure and task

Each participant underwent three sessions in this study. First, they were invited to the lab for a *screening session* 2~3 days prior to the fMRI scanning. In this session, participant (player B) had a chance to jointly earn monetary rewards with another same-sex partner (player A) by estimating the duration of one second in a time-estimation task (Miltner et al. 1997). They were instructed that their performance in each trial of the time-estimation task would be combined with player A's performance. Per trial participant and player A could earn a joint reward (i.e., 10 monetary units [MUs]), only when one or both of them responded correctly. Reponses were considered to be correct when they were within a certain critical time interval, which was adapted dynamically according to participant's performance in previous trials to maintain an average accuracy of about 50% (Boksem et al. 2011). Participants completed 120 trials in total, during which they could only see their own performance on each trial. Participants were not informed about how many trials they responded correctly in the task, but they were told that their performance in each trial would be combined with that of another player (i.e., player A) to determine both reward and contribution. Unbeknownst to participants, however, there was no player A in the experiment. Participants were told that they would not meet with player A and each player would perform the task in an anonymous setting. In rewarding trials (i.e., either one or both players responded correctly), contributions to the income were manipulated according to relative performance in the task, including self-contribution (self-correct, other-incorrect), both-contribution (self-correct, other-correct), and other-contribution (self-incorrect, other-correct). Participant was informed that the money jointly earned with the partner would be distributed in a follow-up fMRI experiment implementing an adapted version of UG.

Second, participant returned the next 2~3 days for the *fMRI scanning session* to complete the UG, in which the joint earnings of each trial (10 MUs) were divided by a random number generator (RNG) between the participant (player B) and player A (Fig. S1). These income distributions were presented to the participant together with the contribution information (i.e., self-contribution [You: √; Other: ×], both-contribution [You: √; Other: √], and other-contribution [You: ×; Other: √]). In response to each income distribution, participant needed to either accept (both got paid accordingly) or reject (neither got paid) it as in the classical UG. Notably, player A was not responsible for the outcome, but his/her payoff depended on participant's decisions on the income distributions. The RNG manipulation allowed us to reasonably present advantageous income distributions to the participant in different contribution conditions. Such adaptations of UG have also been previously verified for similar efficacy as the standard UG (e.g., Civai et al. 2012; McAuliffe et al. 2013; Yu et al. 2014; Blake et al. 2015). Additionally, in free-win trials indicated by the "Lottery" in the screen, the RNG generated random amounts of MUs only to the participant, and the participant could also either accept or reject these outcomes without any impact on player A's payoff. This free-win condition did not involve social comparison and was employed to control for simple reception of monetary rewards (Civai et al. 2012, 2014). Indeed, participants accepted almost all of divisions in the free-win condition. That's being said, however, it should be noted that other potential effects (e.g., endowment effects) might also differentiate the free-win condition from contribution-related conditions.

Participant completed two fMRI runs lasting about 11 min each. Each run consisted of 84 rounds: eight of advantageous inequality (four for 9:1, two for 8:2, and two for 7:3), eight of equality (5:5), and eight of disadvantageous inequality (two for 3:7, two for 2:8, and four for 1:9) for each of the three contribution conditions (self-contribution, both-contribution, other-contribution). Additionally, there were twelve free-win task rounds (two with 9 MUs, one with 8 MUs, one with 7 MUs, four with 5 MUs, one with 3 MUs, one with 2 MUs, and one with 1 MU).

Participant received instructions about the UG and played four rounds of the game to get familiar with the task before entering the scanner. On each round of UG, an income distribution was presented together with context information (contribution or free-win) constantly for 4 s, during which participant had to decide to either accept or reject the distribution of earnings through a response box. Associations between buttons and decisions were counterbalanced across subjects. At the end of each round, an optimized temporal jitter (ranged from 1 to 7 s, average of 4 s) was presented for statistical efficiency (http://www.cabiatl.com/CABI/resources/fmrisim/). Stimulus presentation and behavioral data collection were implemented by Psychtoolbox-3 (http://psychtoolbox.org/).

Finally, in the *post-scan session,* participant reported subjective judgments of unfairness on a 1 to 7 Likert scale for each type of the income distributions. These subjective ratings were not collected for the first 4 participants. Accordingly, data from only 19 (9 females) (mean age ± s.d.: 22.95 ± 2.09) participants were available.

To encourage real decisions from participants, it was emphasized that they would be paid according to their choices in the game, in addition to fixed show-up compensation. However, participants were not informed about the exchange rate between the MUs and monetary payoff, and each participant was paid the same (150, about $25) at the end of experiment (Fehr and Fischbacher 2003; Corradi-Dell'Acqua et al. 2013; Grecucci et al. 2013). Before leaving the laboratory, participants filled debriefing questionnaires designed to examine their beliefs about the experimental setup. No participants expressed doubts about whether payoffs of both their own and player A were dependent on their decisions in the game.

Data acquisition

Imaging was performed on a 3 T Siemens Trio scanner equipped with a 12-channel transmit/receive gradient head coil at BNU's Imaging Center for Brain Research. A T2-weighted gradient-echo echo-planar imaging (EPI) sequence was used to acquire functional images: TR/TE = 2000 ms/30 ms, flip angle = 90°, number of axial slices = 33, slices thickness = 3.5 mm, gap between slices = 0.7 mm, matrix size = 64 × 64, and FOV = 224 mm × 224 mm (voxel size = 3.5 mm × 3.5 mm × 3.5 mm). There were two EPI scan runs, with 336 volumes per run. High-resolution anatomical images covering the entire brain were obtained by a magnetization prepared rapid acquisition with gradient-echo (MPRAGE) sequence: TR/TE = 2530 ms/3.39 ms, flip angle = 7°, number of sagittal slices = 144, slices thickness = 1.33 mm, matrix size = 256 × 256, FOV = 256 mm × 256 mm (voxel size = 1 mm × 1 mm × 1.33 mm).

Statistical analysis

Behavioral data

Decisions and unfairness ratings All participants responded to at least 94.0% of rounds during the game. Trials with no response were excluded from both behavioral and neuroimaging analyses. Participants' decisions (accept/reject) were analyzed with a generalized linear mixed model (GLMM), implemented in R statistical package (version 3.2.3), using the 'lme4' package (Bates et al. 2014). We tested our hypothesis that the modulating effects of contribution would be smaller for advantageous inequality than disadvantageous inequality (i.e., egocentric bias). This was implemented with the contrast of ([self-contribution versus other-contribution]$_{\text{disadvantageous}}$ versus [other-contribution versus self-contribution]$_{\text{advantageous}}$), with the contribution (coded as "mismatch" and "match" to implement the contrast with "match" as the reference level) and outcomes ("disadvantageous" and "advantageous" with "advantageous" as the reference level) as within-subjects factors and participants as a random factor. Please note that a contrast of [self-contribution versus other-contribution] was used for disadvantageous inequality, whereas a contrast of [other-contribution versus self-contribution] was used for advantageous inequality. This was due to the reason that both disadvantageous inequality in self-contribution and advantageous inequality in other-contribution are mismatched between efforts and pay-offs; in contrast, both disadvantageous inequality in other-contribution and advantageous inequality in self-contribution are matched between efforts and pay-offs. In other words, the contrast of interest equals a 2 (outcome: disadvantageous, advantageous) × 2 (contribution: mismatched, matched) interaction. Therefore, the contrast was coded in the GLMM as [mismatch - match]$_{\text{disadvantageous}}$ versus [mismatch - match]$_{\text{advantageous}}$. A similar analysis approach was implemented for unfairness ratings, except that a paired t-test was employed using SPSS 16.0 (IBM, Somers, USA).

In addition, to test the distribution characteristics of the contribution effects in disadvantageous and advantageous inequality, we adopted a bootstrapping procedure. We combined effects of contributions on rejection rates and unfairness ratings of each participant in each inequality condition as a single data point (contribution effects: rejection rates, unfairness ratings). A bootstrapped dataset was created by resampling the data with replacement, keeping the sample size of data as the number of participants; this procedure was repeated for 2000 times to estimate the mean of the population in both disadvantageous and advantageous contexts.

fMRI data: Activation analysis

Neuroimaging data analyses were performed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) implemented in Matlab (version R2014a). Preprocessing of functional data included realignment through rigid-body registration to correct for head motion, slice-timing correction, normalization to MNI space, interpolation of voxel sizes into $2 \times 2 \times 2$ mm^3, spatial smoothing (8-mm full-width/half-maximum kernel), and high-pass filtering ($f > 0.01$ Hz).

A two-level general linear model (GLM) was used to analyze the fMRI data. In the first level analysis, 12 boxcar (4 s) stimulus functions were defined according to different permutations of the 3 types of income distribution (advantageous inequality, equality, and disadvantageous inequality) and 4 types of contextual information (self-contribution, both-contribution, other-contribution, and free-win). To account for different response times across trials, response time was included as an parametric modulator for each regressor as suggested by previous works (Poldrack et al. 2011; Mumford and Poldrack 2014). Furthermore, trials that did not register participant's responses were excluded from data analysis as an extra nuisance regressor. These regressors, being convolutions between respective boxcar stimulus function with the canonical hemodynamic impulse-response function (HRF) (Büchel et al. 1998), were included in the design matrix together with six head movement parameters. The GLM also modeled first-order temporal autocorrelations in the residual to account for temporal autocorrelations in BOLD fMRI signal (Woolrich et al. 2001). In the second (group) level analysis, the influence of egocentric bias was examined by comparing the modulating effects of contribution on neural responses to advantageous inequality and disadvantageous inequality. Similar to behavioral data analyses, this was achieved by the contrast of $(\text{self-contribution versus other-contribution})_{\text{disadvantageous}}$ versus $(\text{other-contribution versus self-contribution})_{\text{advantageous}}$, i.e., a 2 (outcome: disadvantageous, advantageous) × 2 (contribution: mismatched, matched) interaction. For false positive control, we used whole-brain cluster correction (implemented in SPM12) with a cluster-defining threshold of $P < 0.001$ and a Family Wise Error (FWE) corrected threshold of $P < 0.05$ (Woo et al. 2014; Eklund et al. 2016).

fMRI data: Correlation analysis

To identify brain regions associated with the egocentric bias of fairness decision-making, voxel-wise contrast of interest (i.e., $[(\text{self-contribution versus other-contribution})_{\text{disadvantageous}}$ versus $(\text{other-contribution versus self-contribution})_{\text{advantageous}}]$) was correlated with the egocentric bias in individual behavioral measures of rejection rates. Multiple comparisons were corrected with the same approach and thresholds used in the activation analysis.

fMRI data: Connectivity analysis

The correlation analysis identified the thalamus, in which changes in neural responses were correlated with the egocentric bias of fairness decision-making (see also Results section). Therefore, we examined whether the thalamus worked together with other brain regions to underlie the egocentric bias with an analysis of Psychophysiological interaction (PPI) (Friston et al. 1997) using the thalamus as a region of interest (ROI). Specifically, we used the generalized PPI toolbox (http://www.nitrc.org/projects/gppi, version 13.1) (McLaren et al. 2012) with fMRI signal time courses individually extracted from the thalamus as the seeding signals. These seeding signals were then deconvolved with the canonical HRF, resulting in estimates of the underlying neuronal activity (Gitelman et al. 2003). Subsequently, the interactions of these estimated neuronal time-series and vectors representing each of the onsets for each type of income distribution were computed. Lastly, these interaction terms were re-convolved with the HRF and entered into a new GLM along with the vectors for the onsets of each event (i.e., the psychological terms), the original average time-series and nuisance regressors (i.e., 6 movement parameters derived from realignment corrections). Group level analysis of the PPI data were almost identical to that of activation data except the beta values used were derived from the PPI regressors. In the present study, we focused on connections that exhibited a differential contribution effect between advantageous inequality and disadvantageous inequality (i.e., the egocentric bias). Multiple comparisons were corrected with the same approach and thresholds used in the activation analysis.

Meta-analysis on brain imaging studies of fairness-related decision-making

To examine whether the egocentric bias reflects the modulations of neural activity in brain regions consistently involved in fairness-related decision-making, we performed a meta-analysis on brain imaging studies of fairness-related decision-making. Afterwards, we assessed the overlap between regions involved in the egocentric bias and those consistently engaged by fairness-related decision-making. Details on the meta-analysis approach could be obtained in one of our recent publications (Feng et al. 2015). In brief, we built on our previous meta-analysis on the topic by including additional 10 studies through systematic search and selection, resulting in 27 studies (including a total of 309 peak foci) that reported the contrast of "unfair > fairness" (Table 1). A coordinate-based meta-analysis of those studies was conducted by employing the revised ALE algorithm (Eickhoff et al. 2009) with in-house MATLAB scripts. ALE determines the convergence of foci reported from different functional neuroimaging studies with published foci in Talairach or MNI space (Turkeltaub et al. 2002; Laird et al. 2005). The updated meta-analysis identified the consistent involvement of the AI, dACC and dlPFC in fairness perception, which largely replicated our previous findings (Feng et al. 2015).

Results

Behavioral results

The GLMM on participants' decision-making indicated that the modulating effects of contribution were smaller in response to advantageous distributions than to disadvantageous

Table 1 Summary of studies included for the meta-analysis on unfairness in the Ultimate Game

Study	N	Task and contrast	No. of foci
Baumgartner et al. (2011)	32	Responders in UG, unfair>fair	17
Civai et al. (2012)	19	Responders in a modified UG/DG, unequal>equal	12
Corradi-Dell'Acqua et al. (2016)	19	Responders in UG, unfair > fair	21
Farmer et al. (2016)	18	Responders in UG, unfair > fair	6
Fatfouta et al. (2016)	23	Responders in UG, unfair > fair	18
Feng et al. (2016a)	40	Responders in UG, unfair > fair	10
Haruno et al. (2014)	62	Responders in UG, parametric analysis, positive correlation with inequity.	4
Gospic et al. (2011)	17	Responders in UG, unfair>fair	4
Gradin et al. (2015)	25	Responders in UG, unfair>fair	10
Guo et al. (2014)	18	Responders in UG, unfair>fair	10
Guo et al. (2013)	21	Responders in UG, unfair>fair	13
Güroğlu et al. (2011)	68	Responders in UG, unfair>fair	9
Halko et al. (2009)	23	Responders in UG, unfair>fair	22
Harlé and Sanfey (2012)	38	Responders in UG, unfair>fair	12
Hu et al. (2015)	23	Responders in UG, unfair>fair	4
Kirk et al. (2011)	40	Responders in UG, unfair>fair	11
Kirk et al. (2016)	50	Responders in UG, unfair > fair	11
Roalf (2010)	27	Responders in UG, unfair>fair	8
Sanfey et al. (2003)	19	Responders in UG, unfair>fair	17
Servaas et al. (2015)	114	Responders in UG, unfair>fair	32
Verdejo-García et al. (2015a)	19	Responders in UG, unfair>fair	4
Verdejo-García et al. (2015b)	44	Responders in UG, unfair>fair	13
White et al. (2013)	20	Responders in UG, parametric analysis, positive correlation with unfairness level	8
White et al. (2014)	21	Responders in UG, parametric analysis, positive correlation with unfairness level	7
Wu et al. (2014)	18	Responders in UG, parametric analysis, negative correlation with subjective utility	7
Zheng et al. (2014)	25	Responders in UG, unfair>fair	15
Zhou et al. (2014)	28	Responders in UG, unfair>fair	4

N, number of subjects

distributions (parameter estimate = 2.62, s.e. = 1.22, odds ratio = 13.78, z = 2.14, $P < 0.05$) (Fig. 1a). Similarly, the analysis on unfairness ratings revealed that the influence of contribution was smaller on advantageous distributions than disadvantageous distributions ($t_{18} = 2.24$, $P < 0.05$, Cohen's $d = 0.46$) (Fig. 1b). These findings indicated that participants considered less about earning contribution when assessing advantageous distributions than disadvantageous distributions. Notably, the egocentric bias (i.e., the differential effects of contribution) in unfairness ratings showed positive correlations with the egocentric bias in rejection rates (Spearman $\rho = 0.55$, $P < 0.05$) (Fig. 1c).

Behavioral patterns in individual participants were illustrated in Fig. 2 for both rejection rates (Fig. 2a) and unfairness ratings (Fig. 2b). In addition, the bootstrapped data showed a boundary between disadvantageous and advantageous inequality, supporting that the modulating effects of contribution on both rejection rates and unfairness ratings were smaller for advantageous inequality than disadvantageous inequality (Fig. 2c).

fMRI activation results

The contrast of (self-contribution versus other-contribution)$_{\text{disadvantageous}}$ versus (other-contribution versus self-contribution)$_{\text{advantageous}}$ revealed the activations in the following brain regions (Fig. 3a and Table 2): bilateral AI (Fig. 3b, c), dACC (Fig. 3d), and right dlPFC (Fig. 3e) among others ($P <$ 0.05, FWE corrected at the cluster level; Table 2). In accord with behavioral findings, the modulating effects of contribution on the neural responses of these brain regions were smaller for advantageous distributions than disadvantageous distributions. Notably, brain regions identified for the egocentric bias showed substantial overlap with those consistently involved in the fairness-related decision-making (Fig. 4).

fMRI correlation results

Egocentric bias in rejection rates showed a significant and positive correlation with the neural responses to the contrast of interest (i.e., [self-contribution versus other-

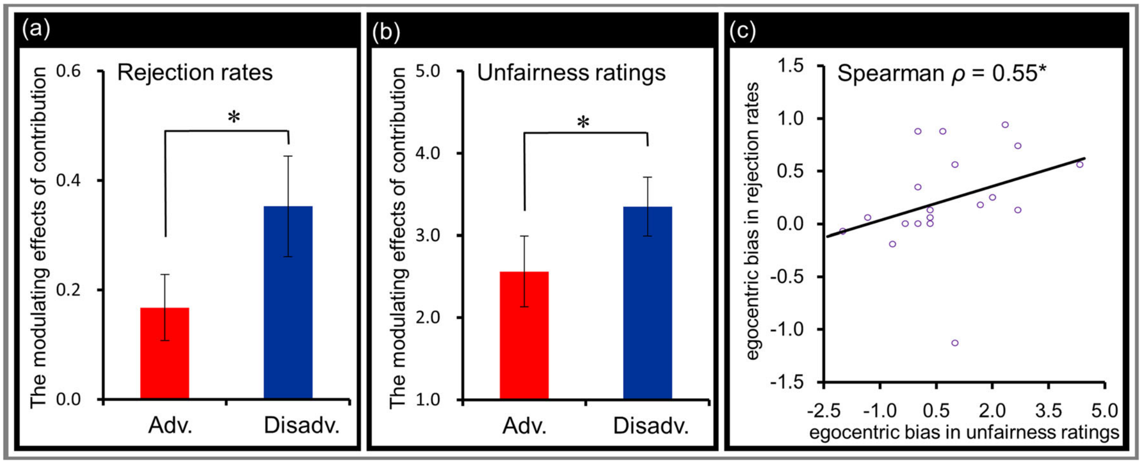

Fig. 1 **Behavioral performance.** The modulating effects of contribution were smaller in the advantageous than disadvantageous distributions for both rejection rates [Advantageous: −0.06~0.94; Disadvantageous: −0.50~1.00] (**a**) and unfairness ratings [Advantageous: −0.67~4.67; Disadvantageous: 0.00~6.00] (**b**). The egocentric bias in rejection rates and unfairness ratings were positively correlated with each other (**c**). *$P < 0.05$. Error bars indicate standard errors. Adv., advantageous inequality; Disadv., disadvantageous inequality. The modulating effects of contribution were defined as differences in behaviors between mismatched and matched contribution conditions, respectively for advantageous (other-contribution minus self-contribution) and advantageous (self-contribution minus other-contribution) outcomes. Larger values of the effect indicate more pronounced unfair feelings to mismatched contribution and outcomes

contribution)$_{\text{disadvantageous}}$ versus (other-contribution versus self-contribution)$_{\text{advantageous}}$]) in the thalamus (x/y/z = 22/2/−2 mm, cluster size = 700; Fig. 5a). That is, the two measurements of egocentric bias, one derived from neuronal level of thalamus activation and the other derived from behavioral level of rejection rate, were positively correlated with each other (Fig. 5b).

PPI results

PPI analysis was performed to assess the effects of egocentric bias on the functional connectivity between the thalamus and other brain regions. The contrast of (self-contribution versus other-contribution)$_{\text{disadvantageous}}$ versus (other-contribution versus self-contribution)$_{\text{advantageous}}$ revealed context-dependent

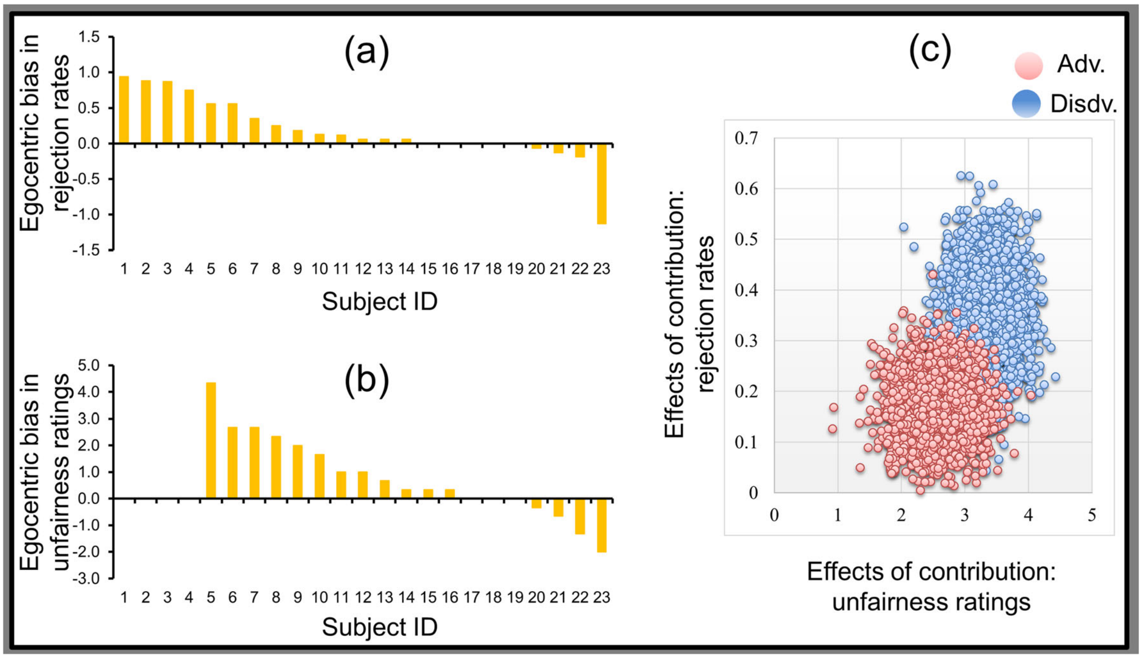

Fig. 2 **Behavioral patterns for individual participants.** The egocentric bias in rejection rates (**a**) and unfairness ratings (**b**) in each participant is illustrated. Bootstrap results of the modulating effects of contribution are illustrated as a function of behavioral measures (rejection rates, unfairness ratings) and inequality (advantageous, disadvantageous) (**c**). The Bootstrap results indicate that the modulating effects of contribution in both rejection rates and unfairness ratings were smaller for advantageous inequality than for disadvantageous inequality. Adv., advantageous inequality; Disadv., disadvantageous inequality. Please note that (i) participants were ordered according to egocentric bias in rejection rates and unfairness ratings respectively in (a) and (b); and (ii) unfairness ratings have missing values from 4 subjects, as shown in (b)

functional connectivity of the thalamus with the following brain regions (Fig. 6a and Table 3): bilateral insula (Fig. 6b, c) and dACC (Fig. 6d) among other brain regions. Notably, the effects of the egocentric bias on the thalamus-left insula connectivity strength (Spearman $\rho = 0.42$, $P < 0.05$) and the thalamus-right insula connectivity strength (Spearman $\rho = 0.45$, $P < 0.05$) showed positive correlations with the egocentric bias in rejection rates.

Discussion

By bridging event-related fMRI, meta-analysis, and context-dependent functional connectivity, the current study examined neural signatures underlying the effects of egocentric bias on fairness decision-making and judgments. We demonstrated an egocentric bias in interpreting contribution during fairness decision-making, such that participants frequently rejected self-contributed (compared to other-contributed) disadvantageous outcomes, but much less so in response to other-contributed (compared to self-contributed) advantageous outcomes, although both involved mismatch between contribution and payoff. This behavioral bias was paralleled with decreased involvement of the AI, dACC, and dlPFC in response to advantageous distributions relative to disadvantageous distributions. Furthermore, the egocentric bias in fairness decision-making was predicted by changes in neural responses of the thalamus, which worked together with fairness-related regions (insula, dACC) to modulate the bias. We interpret these findings as that the egocentric bias of fairness decision-making results from distorted perception of fairness in favor of oneself.

We first demonstrated that participants considered contribution in an egocentric manner, such that participants frequently rejected undeserved disadvantageous outcomes, but they frequently accepted undeserved advantageous outcomes. These findings complement previous observations on the egocentric bias in normative decision-making (Loewenstein et al. 1993; Hoffman et al. 1994; Babcock and Loewenstein 1997; Rodriguez-Lara and Moreno-Garrido 2012) by identifying flexible interpretations of contribution in favor of oneself. The bias could reflect distortions in the perception of fairness,

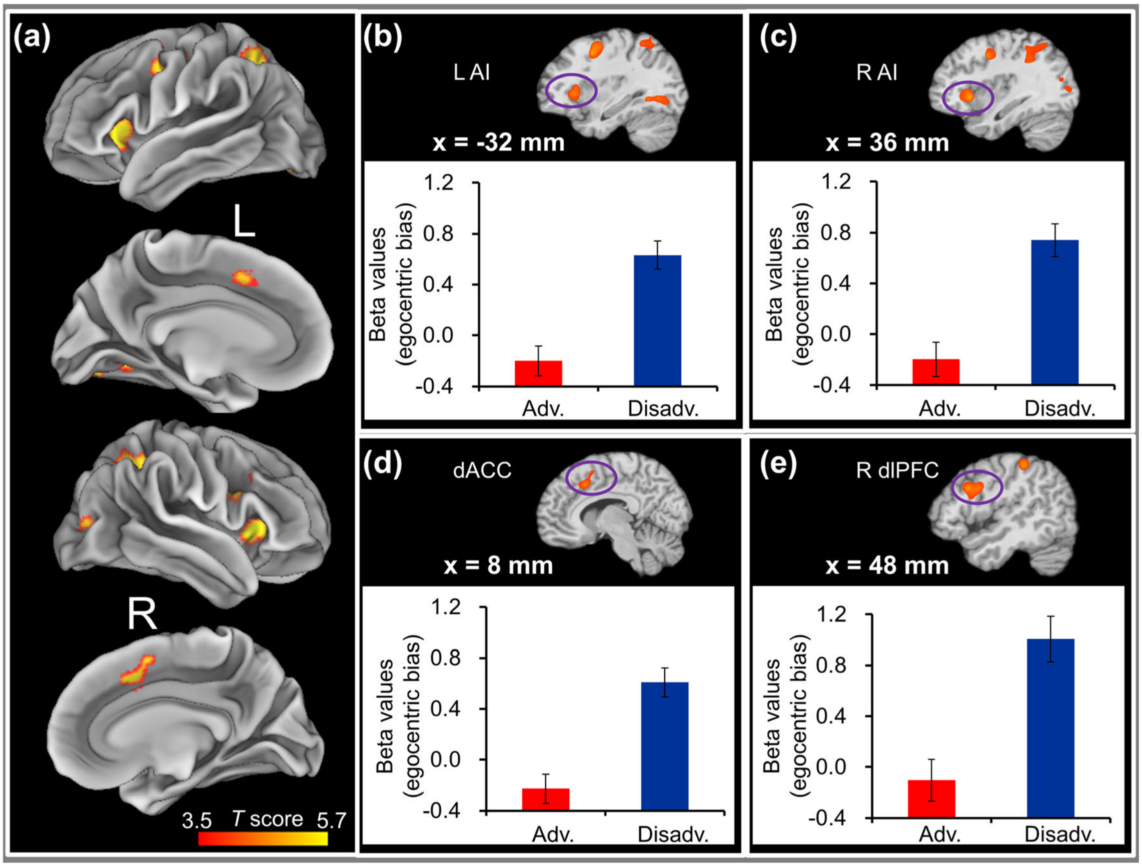

Fig. 3 Brain regions exhibiting weaker modulating effects of contribution in response to advantageous inequality (other-contribution versus self-contribution) than disadvantageous inequality (self-contribution versus other-contribution). The analysis revealed bilateral AI, dACC, and right dlPFC among other brain regions (**a**). Bar plots illustrated parameter estimates of AI (**b & c**), dACC (**d**), and dlPFC (**e**) revealed by the contrast. Images are thresholded at $P < 0.05$ corrected for multiple comparisons at the cluster level in conjunction with voxel-wise $P < 0.001$ (uncorrected; critical cluster size = 200 voxels). L, left; R, right; AI, anterior insula; dACC, dorsal anterior cingulate cortex; dlPFC, dorsolateral prefrontal cortex; Adv., advantageous inequality; Disadv., disadvantageous inequality

Table 2 Brain regions exhibiting differential effects of contribution on neural responses to advantageous inequality compared to disadvantageous inequality

Brain regions	Side	MNI coordination of local maxima (mm)			Local maxima	Cluster size (voxel)
		x	y	z	T	
Contribution effects: adv. > disadv.						
Posterior cingulate cortex	B	−10	−58	22	5.14	389
Contribution effects: adv. < disadv.						
Anterior insula	R	36	20	−4	5.51	285
Anterior insula	L	−28	22	0	5.68	245
Dorsal anterior cingulate gyrus	B	−12	16	42	5.07	211
Middle frontal gyrus	R	50	16	28	5.10	279
Precentral gyrus	R	36	0	40	4.91	214
Precentral gyrus	L	−22	0	46	5.51	609
Inferior parietal lobule	R	24	−54	44	4.57	749
Inferior parietal lobule	L	−28	−56	54	4.65	396
Fusiform	L	−28	−58	−6	5.19	317

Voxel-wise $P < 0.001$ (uncorrected) in conjunction with $P < 0.05$ corrected for multiple comparisons at the cluster level (critical cluster size = 200 voxels), voxel size = $2 \times 2 \times 2$ mm^3 . B, bilateral; R, right; L, left; dis., disadvantageous inequality; adv., advantageous inequality

such that people's intrinsic fairness judgments are biased in favor of oneself (Thompson and Loewenstein 1992; Loewenstein et al. 1993). Additionally, the egocentric bias might be attributed to the self-interest motivation, which drives people to gain immediate rewards and maintain advantage relative to others (Blake et al. 2014; Otto and Bolle 2015). Our brain functional neuroimaging results provide additional line of evidence to dissociate these potential cognitive processes and/or motivations underlying the egocentric bias.

Our neuroimaging findings dovetail with behavioral observations by showing that brain regions important in the detection and resolution of norm violations—the AI, dACC, and dlPFC—exhibited attenuated responses to norm violations for advantageous distributions compared with disadvantageous distributions. Among these regions, the AI and dACC are associated with negative feelings due to fairness-related norm violations (Sanfey et al. 2003; Feng et al. 2016a), whereas the dlPFC is associated with inhibiting self-interested motivation to reject unfair outcomes (Knoch et al. 2006). These regions may work together to facilitate norm enforcement (i.e., rejection) in response to unfairness.

Therefore, the biased neural responses observed in these regions echo the distorted-fairness account, suggesting that the egocentric bias in fairness decision-making reflects the distorted representations of fairness norms (Thompson and Loewenstein 1992; Loewenstein et al. 1993). In line with this idea, our results identified extensive overlap in the AI and dACC between neural responses associated with the egocentric bias and those consistently implicated in fairness perception revealed by meta-analysis.

The distorted-fairness account is further reinforced by our findings that the egocentric bias was predicted by brain activations in the thalamus, a region implicated in encoding aversive feelings during social interactions (White et al. 2013, 2014; Hu et al. 2015). In particular, the thalamus constitutes a pivotal part of the limbic system that is consistently engaged by the processing of negative emotions (Phan et al. 2002). Similarly, this region is involved in the encoding of unfair treatments in UG (Kirk et al. 2011; Hu et al. 2015) and may represent reactive aggressive responses to provocations (White et al. 2014). Therefore, the predicative role of the thalamus for the egocentric bias suggests that the bias in fairness decision-making could result from decreased negative feelings to norm violations in response to advantageous inequality compared with disadvantageous inequality.

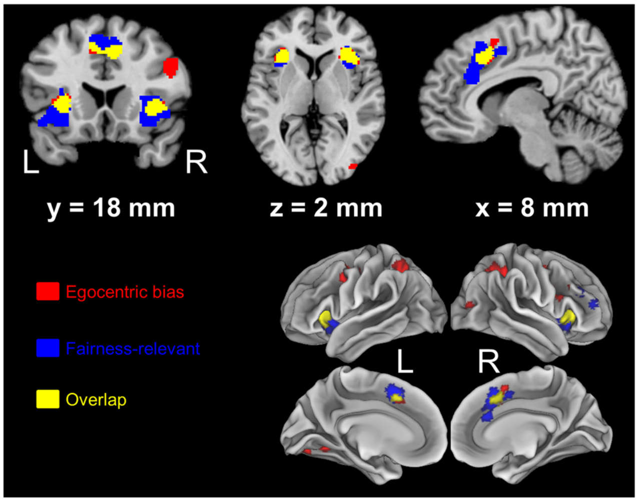

Fig. 4 Overlap in brain regions that were associated with egocentric bias and fairness perception. The overlap was identified in bilateral AI and dACC. L, left; R, right

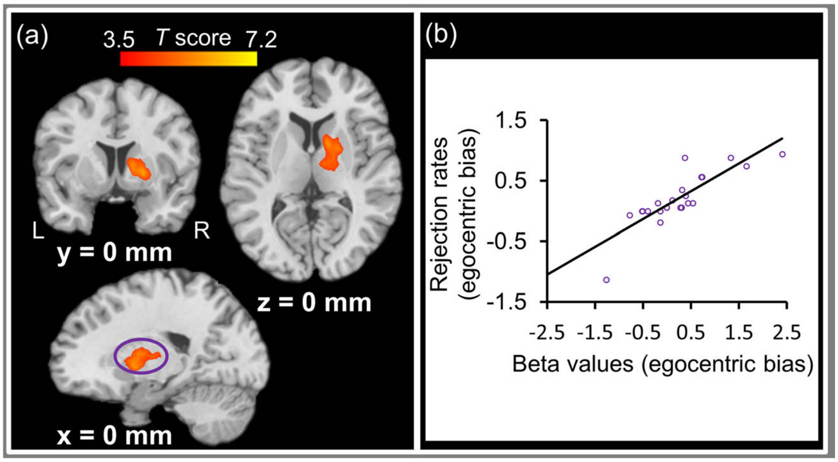

Fig. 5 Brain regions exhibiting positive correlations with rejection rates for the contrast of interest ([other-contribution versus self-contribution)$_{\text{adv.}}$ < (self-contribution versus other-contribution)$_{\text{disadv.}}$]) in the whole-brain analysis. The analysis revealed right thalamus (**a**). Images are thresholded at $P < 0.05$ corrected for multiple comparisons at the cluster level in conjunction with voxel-wise $P < 0.001$ (uncorrected; critical cluster size = 200 voxels). Dot plots illustrated the brain-behavior correlation identified by the analysis **(b)**. Please note that the region was identified by whole-brain analysis,and the dot plots was only for illustration purpose. L, left; R, right

Finally, our functional connectivity findings lend further support to the distorted-fairness account by revealing bias-modulated connectivity between the thalamus and fairness-related regions, including the insula and dACC. Similar to the thalamus, both the insula and dACC have been previously implicated in detecting the presence of norm violation and generating aversive feelings to induce costly punishment of unfairness (Harlé et al. 2012; Corradi-Dell'Acqua et al. 2016). Specifically, a recent neuroimaging study employing multivoxel pattern analysis has identified a domain-general affective processing in the AI and dACC, pointing to a common coding of affective unpleasantness, arousal, and the salience of the experience (Corradi-Dell'Acqua et al. 2016). Taken together, the current findings indicate that the egocentric bias in fairness decision-making is associated with diminished encoding of norm violations, resulting in less aversive responses (i.e., punishment) to norm violations in the advantageous context.

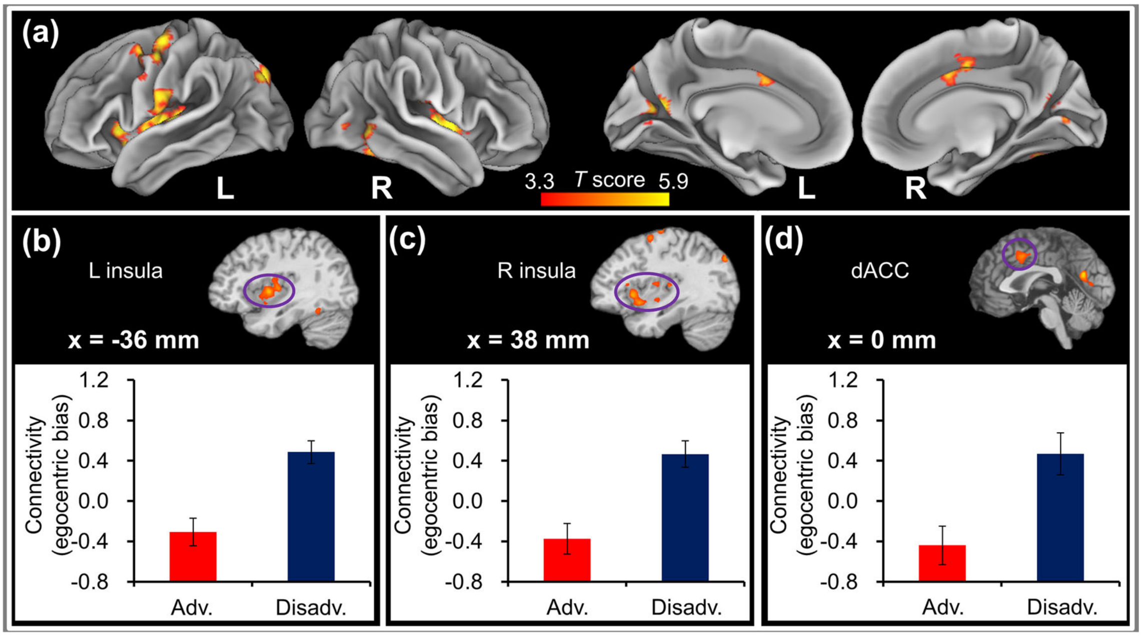

Fig. 6 Brain regions exhibiting functional connectivity with the thalamus as the seed region for the contrast of ([other-contribution versus self-contribution)$_{\text{adv.}}$ < (self-contribution versus other-contribution)$_{\text{disadv.}}$]). The analysis revealed bilateral insula and dACC among other brain regions (**a**). Bar plots illustrated parameter estimates of insula (**b & c**) and dACC (**d**) revealed by the analysis. Images are thresholded at $P < 0.05$ corrected for multiple comparisons at the cluster level in conjunction with voxel-wise $P < 0.001$ (uncorrected; critical cluster size = 168 voxels). L, left; R, right; AI, anterior insula; dACC, dorsal anterior cingulate cortex; Adv., advantageous inequality; Disadv., disadvantageous inequality

Table 3 Brain regions exhibiting stronger effects of contribution on the functional connectivity with thalamus in response to advantageous inequality compared to disadvantageous inequality

Brain regions	Side	MNI coordination of local maxima (mm)			Local maxima	Cluster size (voxel)
		x	y	z	T	
Contribution effects: adv. > disadv.						
Middle temporal gyrus	R	50	−34	−10	5.94	238
Extra-nuclear/caudate	R	28	−18	22	5.08	438
Contribution effects: adv. < disadv.						
Insula	R	42	−12	4	5.39	587
Insula	L	−44	−8	−4	5.67	925
Dorsal anterior cingulate gyrus	B	6	−4	46	5.58	231
Precentral gyrus	L	−44	−14	60	6.11	337
Precuneus	L	−24	−84	36	5.14	266
Calcarine	L	R	−64	18	6.51	187
Inferior temporal gyrus	R	62	−60	−2	7.05	259

Voxel-wise $P < 0.001$ (uncorrected) in conjunction with $P < 0.05$ corrected for multiple comparisons at the cluster level (critical cluster size = 168 voxels), voxel size = $2 \times 2 \times 2$ mm^3 . B, bilateral; R, right; L, left; dis., disadvantageous inequality; adv., advantageous inequality

Notably, our findings extend previous observations on the context-dependent fairness-related behaviors and neural responses by revealing that people interpret social contexts in a manner that favors oneself (i.e., egocentric bias). Prior work has investigated fairness decision-making and its associated brain functions in a variety of social contexts, such as the presence of others (Feng et al. 2016b), membership (Baumgartner et al. 2012; Feng et al. 2016a), social status (Hu et al. 2015), and contributions to the income (Feng et al. 2013; Cappelen et al. 2014; Guo et al. 2014) among others. For instance, the less people contribute to the income, the lower level of brain activity of fairness-related network responds (e.g., the AI, dACC and dlPFC) to disadvantageous divisions (Guo et al. 2014) and the more likely people accept these divisions (Feng et al. 2013; Guo et al. 2014). Although these findings provide important insights on the flexible fairness decision-making and brain functions, however, the current study provides the first and converging evidence to show that social contexts (e.g., contribution) are interpreted in an egocentric manner and that the egocentric bias results from distortions in intrinsic fairness perceptions.

Several limitations of the current study should be noted. First, participants exhibited heterogeneity in egocentric bias (see also Fig. 2a, b), which is in line with previous findings (Bediou and Scherer 2014). For instance, egocentric bias in allocating resources was only found among individualists but not among prosocials (Bediou and Scherer 2014), and it would be both intriguing and important to examine neural correlates mediating those individual differences in egocentric bias. Relatedly, future studies are also needed to examine the egocentric bias among clinical populations as a potential psychological marker. Potential candidates include those patients with social functioning deficit, such as people with narcissistic personality disorder, machiavellianism, and autism spectrum disorder (Silani et al. 2017). Second, the joint earnings were divided by a random device, which is a common approach when advantageous offers are involved (e.g., Civai et al. 2012; McAuliffe et al. 2013; Yu et al. 2014; Blake et al. 2015). This approach is different from the traditional UG with respect to the intentionality of the proposers, which is also an important factor in fairness decision-making (Falk and Fischbacher 2006). Notably, however, previous studies have identified similar behavioral patterns with the current findings even if participants were told that divisions were proposed by another person (Feng et al. 2013; Bediou and Scherer 2014).

Despite these limitations, our findings identified the neural basis underlying the modulation of the egocentric bias in normative decision-making and highlighted the role of the neural network consisting of the AI, dACC and dlPFC that are associated with the representations of fairness norms in this phenomenon. Our findings might have significant implications for understanding real-life bargaining in which fairness is judged in complex contexts and the egocentric bias is common.

Contributions statement CF, XF, LW, and YJL conceived, designed, and performed the experiment. CF, GP and LZ analysed the data. CF, ZL, and YJL wrote the manuscript. LLW, RG, AN, and GD contributed to manuscript revision. YJL provided lab equipment for running the experiment.

Funding This study was funded by the Chinese postdoctoral innovation talent support program (BX201600019), the China Postdoctoral Science Foundation (2017 M610055),the National Natural Science Foundation of China and Senzhen University (31671169, 31500920, 31300869, 31671169, and 201564/000099), the foundation of the National Key Laboratory of Human Factors Engineering (HF2012-K-03), and the Natural Science Foundation of Jiangsu Province of China (BK20130415).

Compliance with ethical standards

Conflict of interest Author Chunliang Feng, Xue Feng, Li Wang, Lili Wang, Ruolei Gu, Aiping Ni, Gopikrishna Deshpande, Zhihao Li, and Yue-Jia Luo declare that they have no conflict of interest.

Ethical approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, and the applicable revisions at the time of the investigation.

Informed consent Informed consent was obtained from all subjects for being included in the study.

References

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *The Journal of Economic Perspectives, 11*(1), 109–126.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014) 'lme4: Linear mixed-effects models using Eigen and S4', R package version, 1(7).

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468–1474.

Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping, 33*(6), 1452–1469.

Bediou, B., & Scherer, K. R. (2014). Egocentric fairness perception: Emotional reactions and individual differences in overt responses. *PLoS One, 9*(2), e88432.

Blake, P., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences, 18*(11), 559–561.

Blake, P., McAuliffe, K., Corbit, J., Callaghan, T., Barry, O., Bowie, A., Kleutsch, L., Kramer, K., Ross, E., & Vongsachang, H. (2015). The ontogeny of fairness in seven societies. *Nature, 528*, 258.

Boksem, M. A., Kostermans, E., & De Cremer, D. (2011). Failing where others have succeeded: Medial frontal negativity tracks failure in a social context. *Psychophysiology, 48*(7), 973–979.

Büchel, C., Holmes, A., Rees, G., & Friston, K. (1998). Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage, 8*(2), 140–148.

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655–661.

Cappelen, A. W., Eichele, T., Hugdahl, K., Specht, K., Sørensen, E. Ø., & Tungodden, B. (2014). Equity theory and fair inequality: A neuroeconomic study. *Proceedings of the National Academy of Sciences, 111*(43), 15368–15372.

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience, 8*(3), 277–284.

Chang, L. J., & Smith, A. (2015). Social emotions and psychological games. *Current Opinion in Behavioral Sciences, 5*, 133–140.

Civai, C., Crescentini, C., Rustichini, A., & Rumiati, R. I. (2012). Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. *NeuroImage, 62*(1), 102–112.

Civai, C., Miniussi, C. & Rumiati, R. I. (2014) Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. *Social Cognitive and Affective Neuroscience*, p. nsu154.

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. *Nature Communications, 7*.

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2013). Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social Cognitive and Affective Neuroscience, 8*(4), 424–431.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping, 30*(9), 2907–2926.

Eklund, A., Nichols, T. E. & Knutsson, H. (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, p. 201602413.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293–315.

Farmer, H., Apps, M., & Tsakiris, M. (2016). Reputation in an economic game modulates premotor cortex activity during action observation. *European Journal of Neuroscience, 44*(5), 2191–2201.

Fatfouta, R., Meshi, D., Merkl, A. & Heekeren, H. R. (2016) Accepting unfairness by a significant other is associated with reduced connectivity between medial prefrontal and dorsal anterior cingulate cortex. *Social Neuroscience*, 1–13.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791.

Feng, C., Luo, Y., Gu, R., Broster, L. S., Shen, X., Tian, T., Luo, Y.-J., & Krueger, F. (2013). The flexible fairness: Equality, earned entitlement, and self-interest. *PLoS One, 8*(9), e73106.

Feng, C., Luo, Y. J., & Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. *Human Brain Mapping, 36*(2), 591–602.

Feng, C., Azarian, B., Ma, Y., Feng, X., Wang, L., Luo, Y. J., & Krueger, F. (2016a). Mortality salience reduces the discrimination between in-group and out-group interactions: A functional MRI investigation using multi-voxel pattern analysis. *Human Brain Mapping, 38*, 1281–1298.

Feng, C., Deshpande, G., Liu, C., Gu, R., Luo, Y. J., & Krueger, F. (2016b). Diffusion of responsibility attenuates altruistic punishment: A functional magnetic resonance imaging effective connectivity study. *Human Brain Mapping, 37*(2), 663–677.

Friston, K. J., Buechel, C., Fink, G., Morris, J., Rolls, E., & Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage, 6*(3), 218–229.

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage, 19*(1), 200–207

Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johannesson, M., & Ingvar, M. (2011). Limbic justice—Amygdala involvement in immediate rejection in the ultimatum game. *PLoS Biology, 9*(5), e1001054.

Gradin, V., Pérez, A., MacFarlane, J., Cavin, I., Waiter, G., Engelmann, J., Dritschel, B., Pomi, A., Matthews, K., & Steele, J. (2015). Abnormal brain responses to social fairness in depression: An fMRI study using the ultimatum game. *Psychological Medicine, 45*(06), 1241–1251.

Grecucci, A., Giorgetta, C., van't Wout, M., Bonini, N., & Sanfey, A. G. (2013). Reappraising the ultimatum: An fMRI study of emotion regulation and decision making. *Cerebral Cortex, 23*(2), 399–410.

Guo, X., Zheng, L., Zhu, L., Li, J., Wang, Q., Dienes, Z., & Yang, Z. (2013). Increased neural responses to unfairness in a loss context. *NeuroImage, 77*, 246–253.

Guo, X., Zheng, L., Cheng, X., Chen, M., Zhu, L., Li, J., Chen, L., & Yang, Z. (2014). Neural responses to unfairness and fairness depend on self-contribution to the income. *Social Cognitive and Affective Neuroscience, 9*(10), 1498–1505.

Güroğlu, B., van den Bos, W., Rombouts, S. A., & Crone, E. A. (2010). Unfair? It depends: neural correlates of fairness in social context. *Social Cognitive and Affective Neuroscience, 5*(4), 414–423.

Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S. A., & Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *NeuroImage, 57*(2), 634–641.

Halko, M.-L., Hlushchuk, Y., Hari, R., & Schürmann, M. (2009). Competing with peers: Mentalizing-related brain activity reflects what is at stake. *NeuroImage, 46*(2), 542–548.

Harlé, K. M., & Sanfey, A. G. (2012). Social economic decision-making across the lifespan: An fMRI investigation. *Neuropsychologia, 50*(7), 1416–1424.

Harlé, K. M., Chang, L. J., van't Wout, M., & Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: A mediating role of the anterior insula. *NeuroImage, 61*(1), 32–40.

Haruno, M., Kimura, M., & Frith, C. D. (2014). Activity in the nucleus Accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience, 26*(8), 1861–1870.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*(3), 346–380.

Hu, J., Blue, P., Yu, H., Gong, X., Xiang, Y., Jiang, C. & Zhou, X. (2015) 'Social status modulates the neural response to unfairness', *Social Cognitive and Affective Neuroscience*, p. nsv086.

Kirk, U., Downar, J., & Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Frontiers in Neuroscience, 5*, 49.

Kirk, U., Gu, X., Sharp, C., Hula, A., Fonagy, P., & Montague, P. R. (2016). Mindfulness training increases cooperative decision making in economic exchanges: Evidence from fMRI. *NeuroImage, 138*, 274–283.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829–832.

Kriss, P. H., Loewenstein, G., Wang, X., & Weber, R. A. (2011). 'Behind the veil of ignorance: Self-serving bias in climate change negotiations. *Judgment and Decision making, 6*(7), 602.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., Turkeltaub, P. E., Kochunov, P., & Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping, 25*(1), 155–164.

Loewenstein, G., Issacharoff, S., Camerer, C., & Babcock, L. (1993). Self-serving assessments of fairness and pretrial bargaining. *The Journal of Legal Studies, 22*(1), 135–159.

McAuliffe, K., Blake, P. R., Kim, G., Wrangham, R. W., & Warneken, F. (2013). Social influences on inequity aversion in children. *PLoS One, 8*(12), e80966.

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage, 61*(4), 1277–1286.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788–798.

Mumford, J. A. & Poldrack, R. A. (2014) 'Adjusting mean activation for reaction time effects in BOLD fMRI', *OHBM poster*, https://ww4.aievolution.com/hbm1401/files/content/abstracts/43589/2053_Mumford.pdf.

Otto, P. E., & Bolle, F. (2015). Exploiting one's power with a guilty conscience: An experimental investigation of self-serving biases. *Journal of Economic Psychology, 51*, 79–89.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16*(2), 331–348.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge University Press.

Roalf, D. R. (2010) 'It's not fair!: behavioral and neural evidence that equity influences social economic decisions in healthy older adults', *Student Scholar Archive*, Paper 516.

Rodriguez-Lara, I., & Moreno-Garrido, L. (2012). Self-interest and fairness: Self-serving choices of justice principles. *Experimental Economics, 15*(1), 158–175.

Rutström, E. E., & Williams, M. B. (2000). 'Entitlements and fairness:: an experimental study of distributive preferences. *Journal of Economic Behavior & Organization, 43*(1), 75–89.

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*(5626), 1755–1758.

Servaas, M. N., Aleman, A., Marsman, J.-B. C., Renken, R. J., Riese, H. & Ormel, J. (2015) Lower dorsal striatum activation in association with neuroticism during the acceptance of unfair offers. *Cognitive, Affective, & Behavioral Neuroscience*, pp. 1–16.

Silani, G., Hartmann, H., Hitz, H., Stepnicka, P. & Lengersdorff, L. (2017) 'Emotional egocentric Bias in autism Spectrum disorder: Behavioral and neurophysiological evidence', Journal of Mental Health Research in Intellectual Disabilities, Taylor & Francis.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19*(4), 339–347.

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes, 51*(2), 176–197.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage, 16*(3), 765–780.

Verdejo-García, A., Verdejo-Román, J., Rio-Valle, J. S., Lacomba, J. A., Lagos, F. M. & Soriano-Mas, C. (2015a) 'Dysfunctional involvement of emotion and reward brain regions on social decision making in excess weight adolescents', Human Brain Mapping, 36(1), pp. 226–237

Verdejo-García, A., Verdejo-Román, J., Albein-Urios, N., Martínez-González, J. M. & Soriano-Mas, C. (2015b) Brain substrates of social decision-making in dual diagnosis: cocaine dependence and personality disorders. Addiction Biology.

White, S. F., Brislin, S. J., Meffert, H., Sinclair, S., & Blair, R. J. R. (2013). Callous-unemotional traits modulate the neural response associated with punishing another individual during social exchange: a preliminary investigation. *Journal of Personality Disorders, 27*(1), 99.

White, S. F., Brislin, S. J., Sinclair, S., & Blair, J. R. (2014). Punishing unfairness: Rewarding or the organization of a reactively aggressive response? *Human Brain Mapping, 35*(5), 2137–2147.

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage, 91*, 412–419.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage, 14*(6), 1370–1386.

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., Jiang, Y. & Zhou, X. (2014) 'Neural basis of increased costly norm enforcement under adversity', *Social Cognitive and Affective Neuroscience*, Epub ahead of print, p. nst187.

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *The Journal of Neuroscience, 33*(3), 1099–1108.

Yu, R., Calder, A. J., & Mobbs, D. (2014). Overlapping and distinct representations of advantageous and disadvantageous inequality. *Human Brain Mapping, 35*(7), 3290–3301.

Zheng, L., Guo, X., Zhu, L., Li, J., Chen, L. & Dienes, Z. (2014) 'Whether Others Were Treated Equally or Not Affects Neural Responses to Unfairness in the Ultimatum Game', *Social Cognitive and Affective Neuroscience*, epub ahead of print, p. nsu071.

Zhou, Y., Wang, Y., Rao, L.-L., Yang, L., & Li, S. (2014). Money talks: Neural substrate of modulation of fairness by monetary incentives. *Frontiers in Behavioral Neuroscience, 8*, 150.