◆ Human Brain Mapping 37:663–677 (2016) ◆

Diffusion of Responsibility Attenuates Altruistic Punishment: A Functional Magnetic Resonance Imaging Effective Connectivity Study

Chunliang Feng,[1,2] Gopikrishna Deshpande,[3,4,5] Chao Liu,[2] Ruolei Gu,[6] Yue-Jia Luo,[2,7]* and Frank Krueger[8,9]

[1]*Institute of Affective and Social Neuroscience, School of Psychology and Sociology, Shenzhen University, Shenzhen, China*
[2]*State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China*
[3]*Department of Electrical and Computer Engineering, Auburn University MRI Research Center, Auburn University, Auburn, Alabama*
[4]*Department of Psychology, Auburn University, Auburn, Alabama*
[5]*labama Advanced Imaging Consortium, Auburn University and University of Alabama Birmingham, Alabama*
[6]*Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*
[7]*Collaborative Innovation Center of Sichuan for Elder Care and Health, Chengdu Medical College, Chengdu, China*
[8]*Molecular Neuroscience Department, George Mason University, Fairfax, Virginia*
[9]*Department of Psychology, George Mason University, Fairfax, Virginia*

Abstract: Humans altruistically punish violators of social norms to enforce cooperation and pro-social behaviors. However, such altruistic behaviors diminish when others are present, due to a diffusion of responsibility. We investigated the neural signatures underlying the modulations of diffusion of responsibility on altruistic punishment, conjoining a third-party punishment task with event-related functional magnetic resonance imaging and multivariate Granger causality mapping. In our study, participants acted as impartial third-party decision-makers and decided how to punish norm violations under two different social contexts: alone (i.e., full responsibility) or in the presence of putative other third-party decision makers (i.e., diffused responsibility). Our behavioral results demonstrated that the diffusion of responsibility served as a mediator of context-dependent punishment. In the presence of putative others, participants who felt less responsible also punished less severely in response to norm violations. Our neural results revealed that underlying this behavioral effect was a network of interconnected brain regions. For unfair relative to fair splits, the presence of others led to attenuated

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: National Natural Science Foundation of China; Contract grant number: 31530031, 81471376; Contract grant sponsor: National Basic Research Program of China (973 Program); Contract grant number: 2014CB744600

*Correspondence to: Yuejia Luo, Institute of Affective and Social Neuroscience, School of Psychology and Sociology, Shenzhen University, Shenzhen, China 518060. E-mail: luoyj@szu.edu.cn

Received for publication 28 April 2015; Revised 16 October 2015; Accepted 6 November 2015.

DOI: 10.1002/hbm.23057
Published online 26 November 2015 in Wiley Online Library (wileyonlinelibrary.com).

responses in brain regions implicated in signaling norm violations (e.g., AI) and to increased responses in brain regions implicated in calculating values of norm violations (e.g., vmPFC, precuneus) and mentalizing about others (dmPFC). The dmPFC acted as the driver of the punishment network, modulating target regions, such as AI, vmPFC, and precuneus, to adjust altruistic punishment behavior. Our results uncovered the neural basis of the influence of diffusion of responsibility on altruistic punishment and highlighted the role of the mentalizing network in this important phenomenon. *Hum Brain Mapp 37:663–677, 2016*.

Key words: altruistic punishment; diffusion of responsibility; mentalizing; functional magnetic resonance imaging (fMRI); Granger Causality mapping

INTRODUCTION

Altruistic punishment of social norm violations is crucial for maintaining widespread cooperation in human societies [Fehr and Fischbacher, 2004a; Fehr and Gächter, 2000]. This type of costly punishment is employed by either directly affected (i.e., second-party punishment) or unaffected (i.e., third-party punishment) individuals, who accept personal costs to reinforce social norms without any overt benefits [Fehr and Fischbacher, 2004b; Henrich et al., 2006; Strobel et al., 2011; Yu et al., 2014]. Recent studies have identified multiple neuropsychological systems that mediate altruistic punishment, including anterior insula (AI), anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (dlPFC), ventromedial prefrontal cortex (vmPFC), precuneus/posterior cingulate cortex (PCC) and other brain regions [Fehr and Camerer, 2007; Feng et al., 2015; Harlé and Sanfey, 2012]. Among these regions, AI, ACC and dlPFC often show stronger responses to norm violations (e.g., unfair splits) than to norm obedience (e.g., fair splits). AI might contribute to signaling norm violations [Chang and Sanfey, 2013; Civai et al., 2012, 2013], whereas ACC and dlPFC reconcile motivational conflicts between altruistic punishment and self-interest [Fehr and Camerer, 2007; Feng et al., 2015; Sanfey et al., 2006]. In contrast, brain regions associated with reward processing (e.g., vmPFC, precuneus/PCC) consistently show stronger responses to fair splits, which is thought to reflect more positive subjective values of fairness [Feng et al., 2015; Xiang et al., 2013].

The neural and behavioral responses to norm violations do not represent simple heuristics, but are sensitive to social contexts [Güroğlu et al., 2010; Hu et al., 2014, 2015; Wu et al., 2014; Yu et al., 2015], supporting the idea that human motivations to enforce social norms are flexible [Chang and Sanfey, 2013]. As core regions underlying altruistic punishment, activations of AI and vmPFC predict amounts of punishment to transgressions by respectively signaling norm deviations [Harlé et al., 2012; Wright et al., 2011] and valuation of social norms [Gu et al., 2015; Xiang et al., 2013] in a context-dependent manner. The effects of social contexts on altruistic punishment often reflect the modulations of other neural networks, such as the mentalizing network consisting of dorsomedial PFC (dmPFC) and temporo-parietal junction [Baumgartner et al., 2012; Güroğlu et al., 2010; Halko et al., 2009] and reappraisal network consisting of ventrolateral PFC (vlPFC) [Grecucci et al., 2013; Tabibnia et al., 2008].

Previous behavioral studies have demonstrated context-dependent altruistic behaviors, such that people's helping behaviors are largely modulated by subjective responsibility [Latané and Nida, 1981]. The constraints of diffusion of responsibility on altruistic behaviors were initially revealed in a landmark study by Darley and Latané [1968], in which the presence of others reduced the likelihood of helping responses. In this study, the authors proposed two key causes of diffusion of responsibility: (i) other people are present and are potentially available to help and (ii) the behaviors of others cannot be closely observed. A plethora of studies have replicated these initial findings in various situations [Fischer et al., 2011; Latané and Nida, 1981]. For instance, the presence of others leads to decreases in the amounts of donations given to an individual [Wegner and Schaefer, 1978] or a charitable organization [Wiesenthal et al., 1983] due to diffusion of responsibility [Mynatt and Sherman, 1975]. Likewise, the amounts of help and tips given out decrease as a function of group size [Freeman et al., 1975; van Bommel et al., 2012].

The behavioral correlates of diffusion of responsibility have been intensely studied in various fields [Fischer et al., 2011; Guerin, 2011]; however, the phenomenon's underlying neural network and its effective connectivity remain unknown. In this study, we combined functional MRI (fMRI) and multivariate Granger causality mapping (GCM) with a third-party punishment task, in which participants acted as impartial third-party decision-makers under two different social contexts: alone (i.e., full responsibility) or in the presence of putative other third-party decision makers (i.e., diffused responsibility). Participants observed how a sum of money was allocated between several pairs of players in a dictator game, where recipients had to accept either fair or unfair splits from dictators [Strobel et al., 2011]. Participants then decided how to punish norm violations committed by the dictator at the expense of their own monetary cost.

In light of previous findings, we hypothesized that in the presence of others, altruistic punishment of norm violations decreases due to a diffusion of responsibility. In addition, decreases in altruistic punishment are associated with attenuated responses in brain regions implicated in signaling norm violations (e.g., AI) and enhanced responses

in brain regions implicated in calculating values of norm violations (e.g., vmPFC, precuneus/PCC) and mentalizing (e.g., dmPFC)/reappraisal (e.g., vlPFC). Finally, mentalizing/reappraisal brain regions (dmPFC/vlPFC) act as the driver of the punishment network and modulate target regions (AI, vmPFC, precuneus/PCC) to adjust punishment behavior.

MATERIALS AND METHODS

Subjects

Twenty-two students (11 females) (mean age $\pm$ SD = 22.9 $\pm$ 1.6 years) participated in the study for monetary compensation. All participants were right-handed, had normal or corrected-to-normal vision, and no neurological or psychiatric history. Written informed consent was 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.

Game Paradigm

Participants acted as third-party decision-makers (player C) and received six monetary units (MUs) for each round of the game. Decision-makers observed how a sum of money (12 MUs) was allocated between several pairs of players (A and B) (i.e., dictator game) [Kahneman et al., 1986]. Participants were told that these persons (player A and B) were participating in a previous study, in which they jointly earned a bonus (12 MUs) by completing a different task. One person from each pair was randomly chosen as player A (dictator) and was asked to allocate the jointly earned money, whereas the other player B (recipient) had to accept A's allocation. Participants were then given a chance to reduce A's payoff as a punishment by altruistically spending their own money: each MU spent reduced three MUs from A's payoff [Bernhard et al., 2006; Fehr and Fischbacher, 2004b]. Participants were instructed that some decisions of A would be only presented to them (alone context); whereas other decisions would be presented to them and four other putative players (C) simultaneously (group context), who were also performing the same task outside the scanner. Under the group condition, participants were told that the MUs spent by the five players (C) would be added to reduce A's payoff. Participants were told that player A might end up with a loss in the case of severe sanctions, which would be compensated by A's show-up fee [see also Fehr and Fischbacher, 2004b]. To encourage real decisions from participants, it was emphasized that MUs were convertible to monetary payoff, and that they would be paid according to their choices in the game, in addition to a fixed show-up compensation. However, participants did not know the exact exchange rate between MUs and monetary payoff, and each participant was paid the same amount of money (¥150 RMB, about $25) at the end of experiment to comply with local ethical guidelines (for details, see also Supporting Information Discussion and Supporting Information Figs. 1 and 2) [Civai et al., 2014; Corradi-Dell'Acqua et al., 2013; Grecucci et al., 2013].

Stimulus presentation and behavioral data collection were implemented by using Psychotoolbox (http://psychtoolbox.org/) [Brainard, 1997; Pelli, 1997]. On each round, a fixation was presented (1 s), followed by the context information (alone context or group context) (1 s) (Fig. 1a). Thereafter, A's allocation was presented constantly for 6 s, during which time participants had to decide how many MUs they were willing to spend to reduce A's payoff using one of the four possible choices: 0, 2, 4, or 6 MUs. Participants made their decisions through a response box, and associations between buttons and decisions were counterbalanced across subjects. After the presentation of A's allocation, an optimized jitter generated by an fMRI simulator software (http://www.cabiatl.com/CABI/resources/fmri-sim/) was presented with minimum of 1 s and average of 4 s. Participants completed two runs each lasting about 10 min (312 scans every 2 s). Each run consisted of 52 rounds: 7 rounds of 12:0 splits, 2 rounds of 11:1, 10:2, 9:3, 8:4, and 7:5 splits, and 9 rounds of 6:6 splits for both alone and group contexts. To mitigate loss of statistical power, splits of 6:6, 7:5, and 8:4 were clustered as fair splits, whereas splits of 9:3, 10:2, 11:1, and 12:0 were clustered as unfair splits. This is according to a recent meta-analysis, indicating that dictators on average generously gave about 30% of their endowment to the recipient [Engel, 2011]. Consequently, equal rounds (n = 26) for each of the following experimental conditions were created: fair splits in the alone context, unfair splits in the alone context, fair splits in the group context, and unfair splits in the group context.

PROCEDURE

Prior to the fMRI session, groups of four to five participants were invited to the lab for a screening session. Participants were informed that the upcoming fMRI session would be conducted in groups of five people, with one participant in the MRI-scanning room and four participants in a room equipped with four computers nearby the MRI-scanning room.

On the day of the fMRI session, the experimental paradigm was explained to the participants, who were then instructed to play four rounds of the game to get familiar with the task. Afterwards, they completed a 10-question multiple-choice quiz designed to assess their understanding of the game paradigm. While participants were prepared for the fMRI session by one experimenter, another experimenter showed up to announce that the other four volunteers participating in the behavioral part of the experiment were ready to begin the experiment. Inside the MRI scanner, participants saw instructions asking all (five) players to press a button in order to begin the experiment, which further

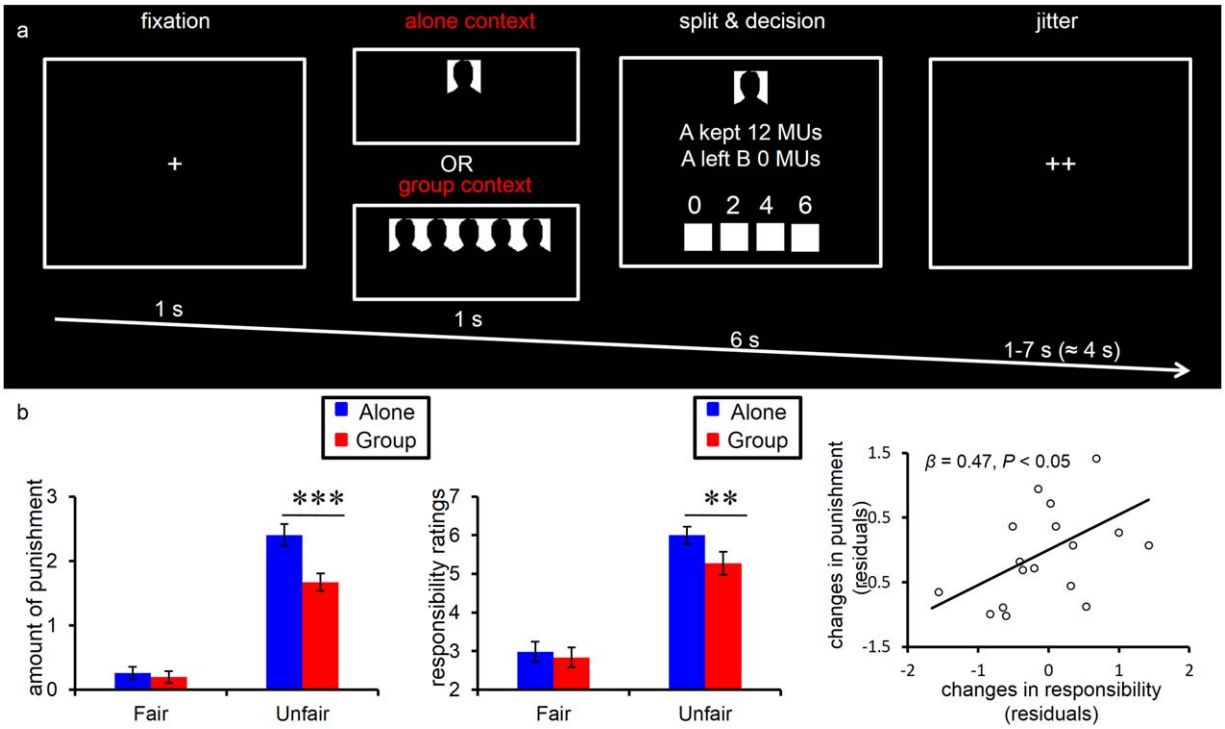

Figure 1.

Task design and behavioral performance. (**a**) Study paradigm. On each round, a fixation was presented and followed by the context information (one person vs. five persons). Then, participants saw A's split and had to decide on how much money (i.e., punishment points) they were willing to spend to reduce A's allocation using one of the four possible choices (monetary units: 0, 2, 4, 6). Finally, an optimized jitter was presented. (**b**) Behavioral results. Participant's punishment to norm violations was attenuated in the group context relative to the alone context. Likewise, participant's sense of responsibility to punish norm violations was attenuated by the group context. Further, participant's sense of responsibility served as a mediator of amounts of behavioral punishment. ***$P < 0.0005$. **$P < 0.005$. Error bars indicate standard error. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

explained that the experiment could only begin after all players had pressed their buttons. In reality, software program triggered the other four button responses automatically. Overall, the aim of these procedures was to increase realism so that participants believed that they were playing with four other people. In a postscan session, participants were asked whether they believed that their payoff would actually be dependent on their decisions in the game, and furthermore, that there were four other players playing with them outside the scanner. Indeed, all participants in the current study believed that their decisions were associated with their payoff and that they were playing with four other players in the game. Therefore, none of the participants were excluded based on their postscan questionnaires.

Participants also completed a survey after the fMRI scanning session. Participants were asked to rate the same splits (fair: 6:6, 7:5, 8:4; unfair: 9:3, 10:2, 11:1, 12:0) observed under both contexts (alone, group) during the experiment on the following seven-point Likert scales: "How much responsibility did you feel to reduce A's money?" (Responsibility: 1 = not at all, 7 = absolutely), "To what extent did you feel that A's allocations were fair?" (Fairness: 1 = absolutely unfair, 7 = absolutely fair), "How excited did you feel?" (Emotional arousal ratings: 1 = very calm, 7 = very excited), and "How pleased did you feel?" (Emotional valence ratings: 1 = very unpleasant, 7 = very pleasant). Due to failures of response, there was incomplete data from four participants (three females); and, therefore, only data from 18 participants were collected for this survey.

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-planar imaging (EPI) sequence was used to acquire functional images: TR/TE = 2,000 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. High-resolution anatomical images covering the entire brain were obtained by applying a magnetization prepared rapid acquisition with gradient-echo (MPRAGE) sequence: TR/TE = 2,530 ms/3.39 ms, flip angle = 7°, number of slices = 144, slices thickness = 1.33 mm, matrix size = 256 × 256, FOV = 256 mm × 256 mm.

Statistical Analysis

Behavioral data

Behavioral data analyses were performed using SPSS 16.0 (IBM, Somers, USA) with a threshold of $P < 0.05$ (two-tailed). Behavioral data were normally distributed (Kolmogorov-Smirnov test) and assumptions for analyses of variance (Bartlett's test) were not violated. To investigate the effects of social context (i.e., diffused responsibility), repeated measure analysis of variances (ANOVAs) on third-party punishment (i.e., amounts of MUs spent), response time, and ratings (i.e., responsibility, fairness, emotional arousal, emotional valence) were applied with Split (fair, unfair) and Context (alone, group) as within-subjects factors.

To further examine the role of subjective responsibility in altruistic punishment, we tested for the mediation effect of subjective responsibility on the difference in amounts of punishment to unfair splits between alone and group contexts, using a regression-based approach proposed for within-subjects designs [Judd et al., 2001]. According to this approach, the mediation effect of subjective responsibility is determined by demonstrating that (i) the subjective responsibility to punish norm violations differs between alone and group contexts, (ii) the amounts of punishment to norm violations differ between alone and group contexts, and (iii) the difference in amounts of punishment between alone and group contexts is predicted by the difference in subjective responsibility (for more details, see also Supporting Information).

FMRI data

Neuroimaging data analyses were performed with SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Preprocessing of functional data included slice-timing correction, realignment through rigid-body registration to correct for head motion, normalization to MNI space, interpolation of voxel sizes to 2 × 2 × 2 mm, smoothing (6-mm full-width/half-maximum kernel), and filtering (high-pass filter set at 128 s).

A two-level general linear model (GLM) was used to analyze the functional data. For the first level, boxcar regressors were defined for each subject and for each epoch of the time course. The regressors modeled the blood-oxygen-level dependent (BOLD) response to the epoch of fixation, context information, and four types of splits (fair splits in the alone and group contexts, unfair splits in the alone and group contexts) with the jitters as implicit baselines. The regressors of each epoch were time-locked to the onset of the epoch with duration from the onset to the offset of the epoch. In particular, the regressors of splits were time-locked to the onset of splits with duration of 6 s [e.g., Harlé et al., 2012]. To control for effects of response time on the BOLD response, response time was added as an additional parametric regressor for each type of split epoch [Poldrack et al., 2011]. This approach is similar to the model that uses orthogonalized response time as the duration of each event [Mumford and Poldrack, 2014]. The six movement parameters of the realignment (three translations, three rotations) were also included in the design matrix as nuisance regressors. Each regressor was convolved with the canonical hemodynamic impulse-response function (HRF) [Büchel et al., 1998] and the resulting GLM was corrected for temporal autocorrelations using a first-order autoregressive model.

For the second level, the interaction between Split (fair, unfair) and Context (alone, group) was assessed by calculating the contrast of [−1 1 1 −1] (fair splits in the alone context, unfair splits in the alone context, fair splits in the group context, unfair splits in the group context). Likewise, main effects of Split and Context were assessed by calculating the contrasts of [−1 1 −1 1] and [−1 −1 1 1], respectively. To correct for multiple comparisons, statistical maps were thresholded at the cluster level with a Monte Carlo simulation-based estimator (3dClustSim: http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html). On the bases of simulations (10,000 iterations) and study parameters (dimensions = 79 × 95 × 68, search volume = 201,139 voxels, voxle size = 2 mm × 2 mm × 2 mm, estimated FWHM = 8.57 mm × 8.63 mm × 8.64 mm, connectivity criterion = surface or edge connected), a family-wise error (FWE) correction at $P < 0.05$ was achieved with a cluster defining threshold of $P < 0.005$ and a cluster size of at least 178 voxels [Lieberman and Cunningham, 2009]. Furthermore, regions of interest (ROIs) were created for brain regions consistently implicated in altruistic punishment (bilateral AI, vmPFC, dlPFC, and dACC) based on a recent meta-analysis [Feng et al., 2015]. These ROIs were defined as spheres with a radius of 10 mm centered at MNI coordinates of $x/y/z = -30/24/2$ mm (left AI), 38/20/0 mm (right AI), 6/46/−12 mm (vmPFC), −30/38/30 mm (left dlPFC), 40/36/26 mm (right dlPFC) and −4/16/48 mm (dACC). For this small volume of 3,090 voxels (including all ROIs), a cluster size greater than 38 voxels was associated with FWE-corrected $P < 0.05$ [see also Corradi-Dell'Acqua et al., 2013; Sebastian et al., 2012; Wu et al., 2014]. Note that clusters that survived cluster-level correction for multiple comparisons either over the whole brain or over the small volume were reported.

To visualize activation patterns for brain regions identified in Split × Context interaction [Poldrack, 2007; Poldrack et al., 2011; Vul and Kanwisher, 2010], parameter estimates of each activated brain region were extracted for all experimental conditions using SPM Rex toolbox (https://www.nitrc.org/projects/rex/). To avoid circularity, no further statistical analyses were performed on these extracted parameter estimates [Kriegeskorte et al., 2009; Vul et al., 2009].

Effective connectivity analysis

A data-driven effective connectivity analysis using multivariate Granger causality mapping (GCM) was implemented [Abler et al., 2006; Grant et al., 2015; Hutcheson et al., 2015; Roebroeck et al., 2005; Wheelock et al., 2014]. This approach provides the ability to assess directional influences among simultaneously recorded BOLD time series in the absence of an a priori model of brain connectivity as required by traditional hypothesis-driven effective connectivity analysis methods such as dynamic causal modeling (DCM) [Friston et al., 2003; Lohmann et al., 2012]. The Granger causality concept draws on the principle of temporal predictability, which assumes that if the current temporal progression of brain activity in one brain region allows the prediction of future temporal progression of activity in another brain region, then the first brain region is assumed to have a causal influence on the second brain region [Granger, 1969]. In recent years, accumulating fMRI studies have employed GCM to reveal effective connectivity among brain regions in various tasks [Friston et al., 2013; Grant et al., 2014; Kapogiannis et al., 2014; Krueger et al., 2011; Lacey et al., 2014; Uddin et al., 2014]. The application of GCM to fMRI data remains debated because of slower temporal sampling relative to faster neuronal processes and non-neural variability of the hemodynamic response. However, recent theoretical developments, simulations, and experimental studies have demonstrated its utility and validity [Davey et al., 2013; Katwal et al., 2013]. For example, using a simple auditory-motor paradigm, Abler et al. [2006] demonstrated that Granger causality can correctly estimate expected causal influences from the auditory cortex to the motor cortex even with fMRI data acquired with long TRs (2,440 ms in their study). Moreover, it has been demonstrated that effective connectivity inferred from GCM and DCM applied to deconvolved fMRI data agreed with those obtained from intracerebral EEG, indicating convergence of evidence from these different methods [David et al., 2008].

For this study, only those regions that survived the fMRI analysis threshold for the interaction effect of Split × Context were selected as ROIs for the subsequent multivariate GCM analysis. For those ROIs, the time series of the BOLD signal intensities were extracted and averaged across voxels and then normalized across participants per run and used for subsequent analysis. Because spatial variability of the BOLD response arising from vascular sources can confound Granger causality obtained from raw fMRI time series [Deshpande et al., 2010], the BOLD time series were subjected to blind hemodynamic deconvolution. The resulting latent neural signals were entered into a first order dynamic multivariate autoregressive (dMVAR) model for assessing directed interactions between multiple nodes as a function of time [Grant et al., 2015; Hutcheson et al., 2015; Wheelock et al., 2014] while factoring out influences mediated indirectly in the set of ROIs selected [Deshpande et al., 2008, 2009; Stilla et al., 2007]. A first order model was applied because of the interest in causal influences arising from neural delays, which are less than a TR [Deshpande et al., 2013]. Blind hemodynamic deconvolution of BOLD signals was performed using a Cubature Kalman filter which has been shown to be extremely efficient for jointly estimating latent neural signals and spatially variable HRFs [Havlicek et al., 2011]. In addition, recent research has shown that this model is not susceptible to over-fitting and produces estimates which are comparable to nonparametric methods [Sreenivasan et al., 2015]. Hemodynamic deconvolution removes the intersubject and inter-regional variability of the HRF [Handwerker et al., 2004] as well as the smoothing effect of the HRF and, therefore, increases the effective temporal resolution of the signal. Further, the dMVAR model's coefficients were allowed to vary as a function of time to obtain condition-specific connectivity values [Sathian et al., 2013]. Boxcar functions corresponding to the split epoch of conditions related to norm violations were used to extract connectivity values corresponding to each context (alone, group). The resulting condition-specific path weights were populated for each condition and paired *t*-tests were performed to assess condition-specific modulations of connectivity in response to norm violations (unfair splits: group context > alone context and alone context > group context). Significant effective connectivity paths were identified using the false discovery rate (FDR, $q(\text{FDR}) < 0.05$) correction for multiple comparisons [Benjamini and Hochberg, 1995].

The BrainNet Viewer toolbox, a graph-theoretical network visualization toolbox under MATLAB (www.mathworks.com), was used to display the effective connectivity networks (i.e., nodes, edges) on a brain surface [Xia et al., 2013].

Bivariate (Spearman ρ) correlations were computed to determine associations among behavioral (decisions, ratings), fMRI (BOLD signal changes) and effective connectivity (path weights) measures.

RESULTS

Behavioral Results

The ANOVA on amounts of punishment revealed significant main effects of Split ($F_{(1,21)} = 118.75$, $P < 0.0005$) and Context ($F_{(1,21)} = 21.55$, $P < 0.0005$), indicating that participants gave stronger punishment in response to unfair splits (norm violations) than to fair splits and in the alone context than in the group context. A significant interaction effect of Split × Context was observed ($F_{(1,21)} = 15.79$, $P < 0.005$) and post hoc comparisons revealed that unfair splits were punished more strongly in the alone context than in the group context ($t_{21} = -4.35$, $P < 0.0005$, Fig. 1b).

The ANOVA on response time revealed significant main effects of Split ($F_{(1,21)} = 114.83$, $P < 0.0005$) and Context ($F_{(1,21)} = 6.11$, $P < 0.05$), indicating that participants spent more time responding to unfair splits (norm violations) than to fair splits and to the alone context than to the group

context. The interaction effect of Split × Context was not significant ($F_{(1,21)} = 0.64$, $P > 0.05$) (Supporting Information Fig. 3).

The ANOVA on responsibility ratings revealed significant main effects of Split ($F_{(1,17)} = 116.83$, $P < 0.0005$) and Context ($F_{(1,17)} = 10.07$, $P < 0.001$), indicating that participants felt more responsible for unfair splits than for fair splits and for splits in the alone context than in the group context. A significant interaction effect of Split × Context was observed ($F_{(1,17)} = 4.94$, $P < 0.05$), demonstrating that participants felt more responsible for punishing norm violations in the alone context compared with the group context ($t_{17} = -3.15$, $P < 0.005$, Fig. 1b).

The mediation analysis revealed that the difference in subjective responsibility to punish norm violations between alone and group contexts was a significant predictor of the difference in amounts of punishment ($\beta = 0.47$, $t_{17} = 2.27$, $P < 0.05$) (Fig. 1c) (for details, see also Supporting Information). Together with the effects of context on both amounts of punishment and subjective responsibility, all three requirements associated with the mediation analysis were met, indicating that subjective responsibility was a mediator of context-dependent amounts of punishment to norm violations.

The ANOVAs on ratings of fairness and emotional feelings (arousal, valence) yielded only a significant main effect of Split (fairness: $F_{(1,17)} = 68.33$, $P < 0.0005$; emotional arousal: $F_{(1,17)} = 20.76$, $P < 0.0005$, and emotional valence: $F_{(1,17)} = 20.67$, $P < 0.0005$), demonstrating that participants' impressions of fairness and feelings of pleasantness were lower for unfair splits than for fair splits, whereas participants felt more arousing in response to unfair splits than to fair splits. However, there was neither a significant main effect of Context (fairness: $F_{(1,17)} = 0.02$, $P > 0.05$; emotional arousal: $F_{(1,17)} = 0.52$, $P > 0.05$, and emotional valence: $F_{(1,17)} = 0.001$, $P > 0.05$) nor a significant interaction effect of Split × Context (fairness: $F_{(1,17)} = 0.03$, $P > 0.05$; emotional arousal: $F_{(1,17)} = 0.01$, $P > 0.05$, and emotional valence: $F_{(1,17)} = 0.05$, $P > 0.05$) (Supporting Information Fig. 3).

fMRI Results

Main effect of split

Unfair splits compared with fair splits activated bilateral AI, dlPFC, ACC, putamen, caudate, middle temporal gyrus, inferior parietal lobule and cerebellum (a more stringent voxel-wise threshold of $P < 0.001$ along with the cluster threshold of $P < 0.05$ FWE corrected was employed to better localize activations) (Fig. 2a,b and Table I). In contrast, fair splits compared with unfair splits activated the bilateral ventromedial prefrontal cortex, precuneus/PCC, parahippocampus gyrus, left inferior prefrontal cortex, right posterior insula and supramarginal gyrus (a more stringent voxel-wise threshold of $P < 0.001$ along with the cluster threshold of $P < 0.05$ FWE corrected was employed to better localize activations) (Fig. 2a,b and Table I).

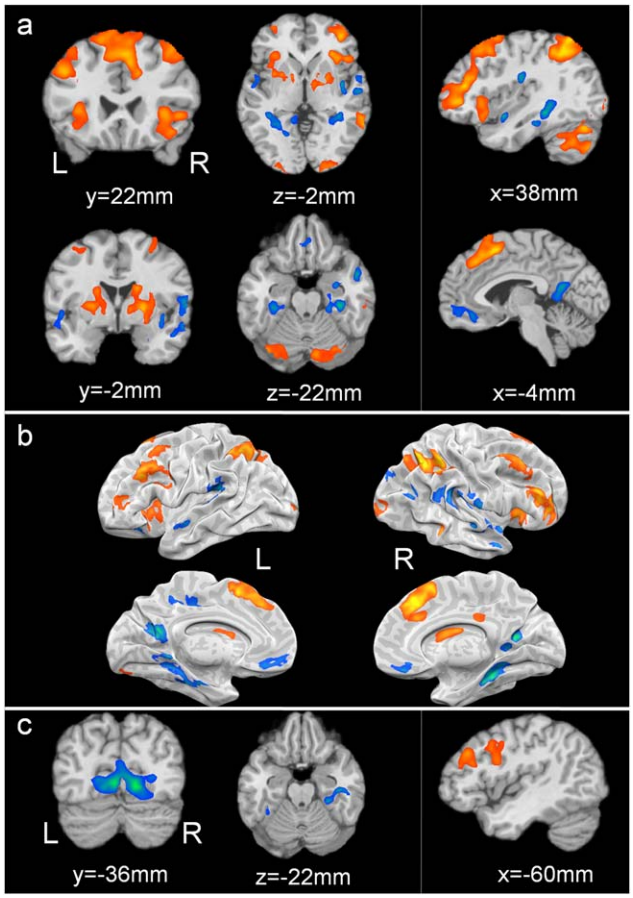

Figure 2.

Influence of the main effects of Split or Context on brain activity. (**a**) and (**b**) Influence of the main effect of Split. Unfair splits compared with fair splits activated stronger activation in bilateral anterior insula, dorsolateral prefrontal cotex, anterior cingulate cortex, putamen, caudate, middle temporal gyrus, inferior parietal lobule and cerebellum (orange). Fair splits compared with unfair splits activated the bilateral ventromedial prefrontal cortex, posterior cingulate cortex, parahippocampus gyrus, left inferior prefrontal cortex, right posterior insula and supramarginal gyrus (blue). (**c**) Influence of the main effect of Context. Activation of bilateral paracentral lobule, left temporal lobe, fusiform, right lingual gyrus and cuneus was larger in the group context than the alone context (blue). The alone context compared with the group context activated the left middle and inferior frontal gyri (orange). L: left; R: right. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Main effect of context

Neural responses of bilateral paracentral lobule, left temporal lobe, fusiform, right lingual gyrus and cuneus were higher in the group context than in the alone context (Fig. 2c and Table I); whereas the alone context compared

TABLE I. Brain regions associated with main effects and interactions

Brain regions	MNI coordination of local maxima (mm)			Local maxima	Cluster size (voxel)
	x	*y*	*z*	*T*	
Main effect of split					
Unfair > fair					
L anterior insula	−30	18	8	6.37	1,180
L putamen	−22	−8	6	6.13	
R inferior frontal gyrus	30	28	−2	5.69	456
R anterior insula	36	22	0	6.61	
R dorsal anterior cingulate cortex	6	26	44	9.10	2,342
R supplementary motor area	2	16	52	7.22	
R middle frontal gyrus	36	38	14	6.82	2,449
R superior frontal gyrus	22	50	32	4.78	187
L middle frontal gyrus	−28	40	14	6.01	649
L middle frontal gyrus	−44	24	34	6.52	1,547
L inferior parietal lobule	−46	−46	50	7.65	1,991
R inferior parietal lobule	46	−50	52	8.57	1,806
R lateral globus pallidus	22	−4	2	7.15	960
R caudate	18	−8	22	5.94	
R middle temporal gyrus	64	−52	−4	6.58	287
L middle temporal gyrus	−54	−54	−14	4.57	91
R middle occipital gyrus	26	−98	−2	4.82	439
L cerebellum	−44	−60	−36	8.10	1,950
R cerebellum	38	−74	−26	6.39	1,631
Fair > unfair					
R ventromedial prefrontal cortex	4	36	−16	5.80	495
L ventromedial prefrontal cortex	−8	36	−16	5.48	
L inferior frontal gyrus	−30	32	−12	6.70	108
R supramarginal gyrus	50	−40	32	6.85	1,163
L temporal lobe	−38	−40	−12	7.96	1,564
L parahippocampa gyrus	−22	−20	−26	6.60	
L superior temporal gyrus	−46	−36	20	7.49	558
R fusiform/parahippocampa gyrus	28	−34	−16	6.88	830
R middle temporal gyrus	52	4	−22	5.52	309
R middle temporal gyrus	46	−66	24	5.53	130
R superior temporal gyrus	−52	2	−8	5.70	231
R precuneus	22	−56	18	7.14	378
L cuneus	22	−92	34	5.15	125
R superior occipital lobe	28	−90	34	6.37	158
R posterior insula	40	2	−14	6.62	305
L precuneus/posterior cingulate cortex	−12	−26	40	5.28	301
Main effect of context					
Alone > group					
L middle frontal gyrus	−44	32	30	4.58	304
L inferior frontal gyrus	−42	28	22	4.47	
L middle frontal gyrus	−50	12	42	5.93	360
Group > alone					
R lingual gyrus	14	−76	−6	6.42	2,376
R cuneus	4	−80	4	6.16	
L temporal lobe	−32	−58	0	5.01	214
L fusiform	−34	−52	−12	4.69	
R paracentral lobule	10	−34	72	4.63	626
L paracentral lobule	−4	−30	66	4.12	
Split × context interaction					
Alone [unfair − fair]> Group [unfair − fair]					
L anterior insula	−34	16	6	5.51	61[a]
R anterior insula	32	24	6	3.91	57[a]

TABLE I. (continued).

Brain regions	MNI coordination of local maxima (mm)			Local maxima	Cluster size (voxel)
	x	*y*	*z*	*T*	
Group [unfair − fair] > alone [unfair − fair]					
L ventromedial prefrontal cortex	4	52	−8	5.93	537
R ventromedial prefrontal cortex	−4	40	−16	4.08	
R precuneus/posterior cingulate cortex	10	−72	20	4.49	626
L precuneus/posterior cingulate cortex	−2	−62	16	3.91	
R dorsomedial prefrontal cortex	8	50	46	3.94	194

All clusters survived correction for multiple comparisons at the cluster level. L, left; R, right.
[a]$P < 0.05$, corrected for the small volume.

with the group context activated the left middle and inferior frontal gyri (Fig. 2c and Table I).

Interaction effect of split × context

The interaction effect revealed changes in BOLD responses among the following regions (Table I): bilateral AI (Fig. 3a,b), vmPFC (Fig. 4a,b), precuneus (Fig. 4a,b) and right dmPFC (Fig. 4a,b). Among these regions, a positive correlation was found between right AI activation and amounts of punishment to unfair splits regarding differences between the alone context and the group context (Spearman $\rho = 0.52$, $P < 0.05$, Fig. 3c). Further, a negative correlation was found between neural responses of vmPFC and subjective responsibility to punish norm violations regarding differences between the alone

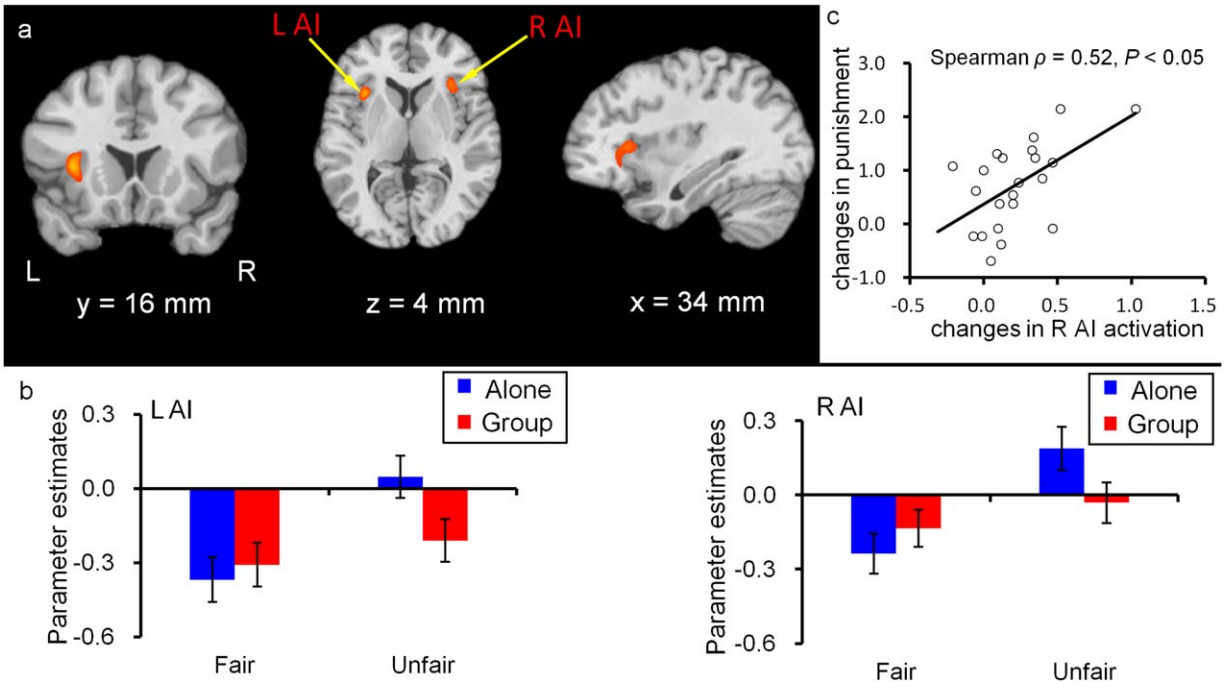

Figure 3.

Influence of the presence of others on the activity of bilateral AI in response to unfair compared with fair splits. (**a**) Activation in bilateral AI revealed by the interaction of Split and Context. (**b**) Parameter estimates of these brain regions as a function of Split and Context. Error bars indicate standard errors that are used for visualization purposes only. No further post hoc comparisons were performed on these extracted data, as this would be considered double dipping. (**c**) Correlation between R AI activation and amounts of punishment to unfair splits regarding differences between alone and group contexts. The correlation remains significant after deleting the outlier (i.e., the data point with the largest changes in R AI activation). L, left; R, right; AI, anterior insula. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

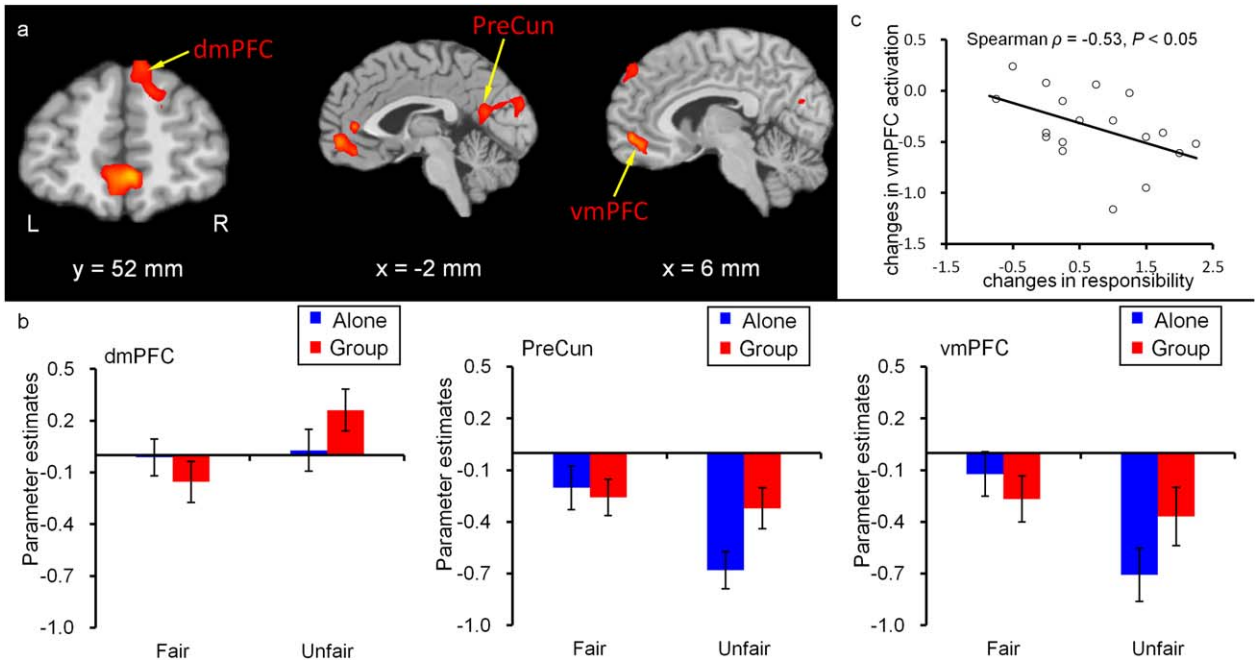

Figure 4.

Influence of the presence of others on the activity of dmPFC, PreCun, and vmPFC in response to unfair splits compared with fair splits. (**a**) Activation in left dmPFC, PreCun, and vmPFC revealed by the interaction of Split and Context. (**b**) Parameter estimates of these brain regions as a function of Split and Context. Error bars indicate standard errors that are used for visualization purposes only. No further post hoc comparisons were performed on these extracted data, as this would be considered double dipping. (**c**) Correlation between vmPFC activation and subjective responsibility to punish unfair splits regarding differences between alone and group contexts. L, left; R, right; dmPFC, dorsomedial prefrontal cortex; PreCun, Precuneus; vmPFC, ventromedial prefrontal cortex. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

context and the group context (Spearman $\rho = -0.53$, $P < 0.05$, Fig. 4c).

EFFECTIVE CONNECTIVITY RESULTS

The multivariate GCM analysis was only performed for those regions that survived the statistical threshold for the Split × Context interaction. Granger causality weights for paths connecting those ROIs were populated for the unfair splits and revealed specific connectivity patterns in the alone context compared with the group context (i.e., unfair splits: alone context > group context) (Table II): The left AI acted as the driver of the network and sent outputs to all other regions of the network (right AI, vmPFC, dmPFC and precuneus) (Fig. 5). In addition, the GCM analysis revealed specific connectivity patterns in the group context compared with the alone context (i.e., unfair splits: group context > alone context) (Table II): The dmPFC acted as the driver of the network and was the only region that sent outputs to all other regions of the network (bilateral AI, vmPFC and precuneus) (Fig. 5).

DISCUSSION

Combining event-related fMRI with multivariate GCM, our study explored the neural signatures underlying the modulations of diffusion of responsibility on altruistic punishment. We demonstrated that impartial third-party decision-makers punished less severely in response to norm violations when in the presence of other punishers, an effect mediated by diffusion of responsibility. Underlying these behavioral effects were neural networks implicated in altruistic punishment (i.e., AI, vmPFC, and precuneus) and mentalizing (i.e., dmPFC). In particular, BOLD responses of bilateral AI to norm violations (compared with fair splits) were lower in the presence of others compared with being alone. In addition, neural responses of vmPFC and precuneus to unfair relative to fair splits were less attenuated in the presence of others compared with being alone. Moreover, the presence of others enhanced neural responses to norm violations in dmPFC. These brain regions were interconnected as a network in the presence of others, such that dmPFC modulated BOLD responses of all other brain regions that were associated with reductions in punishment and subjective responsibility.

TABLE II. Path weights (arbitrary units) from multivariate Granger causality analyses in the alone (top) and group (bottom) contexts in response to norm violations (unfair splits: alone context vs. group context)[a]

Source	Target	Path weights		*T* values	*P*
		Alone contex	Group context		
Alone context > group context					
L AI	PreCun	8.39	2.61	3.85	1.23×10^{-4}
	vmPFC	9.00	2.82	4.10	4.25×10^{-5}
	dmPFC	7.64	1.75	4.01	6.23×10^{-5}
	R AI	8.63	2.43	3.96	7.85×10^{-5}
Group context > alone context					
dmPFC	PreCun	2.32	7.46	2.91	3.61×10^{-3}
	vmPFC	2.02	7.36	3.08	2.09×10^{-3}
	L AI	1.62	7.51	3.42	6.33×10^{-4}
	R AI	0.5	5.17	2.72	6.67×10^{-3}

[a]Positive path weights indicate that BOLD signal changes in the sources and target ROI were in the same direction, whereas negative path weights indicate that BOLD signal changes were in opposite directions in the source and target ROIs. All significant connectivity paths are displayed in the table.
L, left; R, right; AI, anterior insula; PreCun, precuneus; vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex.

We first replicated previous findings on the behavioral and neural correlates of altruistic punishment. Participants in the role of third-party decision-makers punished norm violations at the expense of perceived personal costs. This act of strong reciprocity is considered to be a hallmark of human civilization that allows reinforcement of social

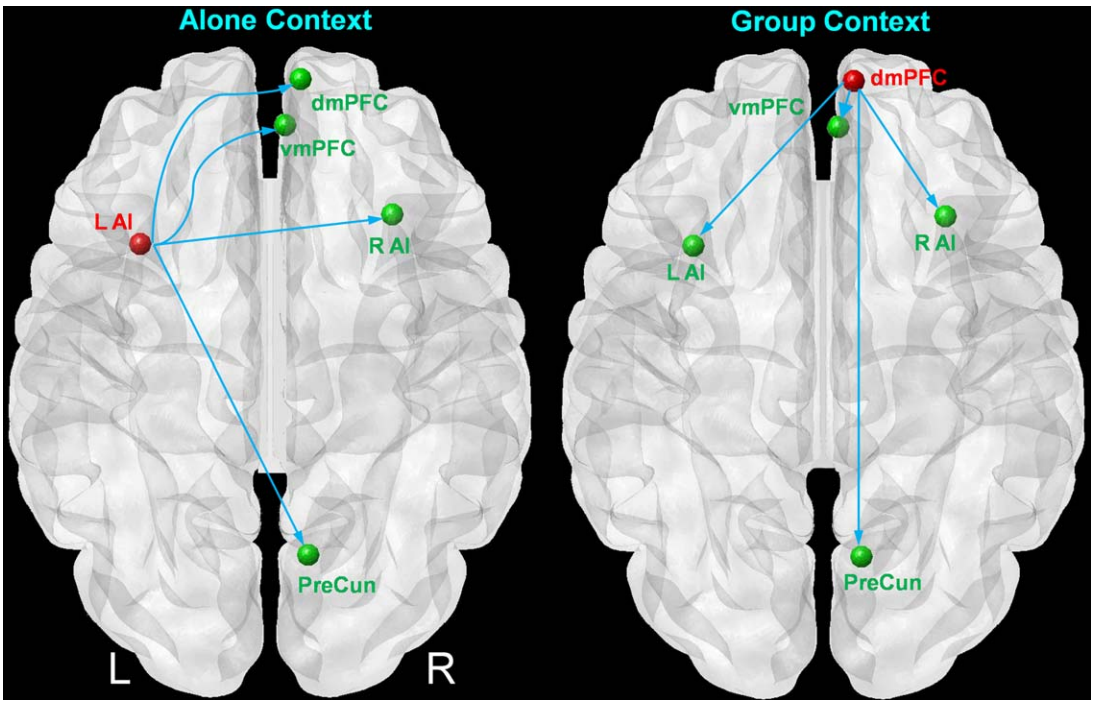

Figure 5.

Connectivity networks revealed in the alone and group contexts in response to norm violations. In the alone context (unfair splits: alone context > group context), left AI acted as the driver of the network and sent outputs to all other regions of the network (precuneus, vmPFC, dmPFC, and right AI). In the group context (unfair splits: group context > alone context), dmPFC acted as the driver of the network and sent outputs to all other regions of the network (precuneus, vmPFC, and bilateral AI). L, left; R, right; AI, anterior insula; PreCun, precuneus; vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

norms to sustain cooperation among individuals of genetically heterogeneous populations [Fehr and Fischbacher, 2004b; Henrich et al., 2006]. Participants' altruistic punishment engaged several brain regions, including AI, vmPFC, precuneus, dlPFC, ACC, and dorsal striatum, presumably reflecting dynamic cognitive-affective-motivational processes that drive this pro-social behavior [Feng et al., 2015; Sanfey and Chang, 2008; Strobel et al., 2011]. As a core region underlying altruistic punishment, AI is thought to detect and mark norm violations (i.e., salient events) for additional processing and to initiate cognitive-affective-motivational processes implemented by other brain regions [Menon and Uddin, 2010]. Our results supported this hypothesis on AI function by demonstrating that left AI acted as the driver of the network to modulate activity of other brain regions at baseline in the alone context.

We next studied the neural basis underlying the modulations of responsibility diffusion on altruistic punishment. Although people's sense of fairness, emotional arousal and valence for norm violations were not affected, diffusion of responsibility was a mediator of context-dependent punishment to norm violations such that people who felt less responsible in response to norm violations punished less severely. Our results confirm diffusion of responsibility as a key psychological mechanism that attenuates altruistic behaviors in the presence of others [Hutcheson et al., 2015; Mynatt and Sherman, 1975; Rosenblatt et al., 1989]. On the neural level, diffusion of responsibility was accompanied by altered neural responses of a network of interconnected brain regions—dmPFC, bilateral AI, vmPFC, and precuneus—enabling people to make flexible context-dependent punishment decisions [Baumgartner et al., 2012, 2013]. In the presence of others as compared with being alone, regions important in detecting norm violations (i.e., AI) showed less pronounced responses to norm violations, and regions associated with encoding of subjective values (i.e., vmPFC and precuneus) showed less attenuation in BOLD responses to norm violations. Notably, the presence of others enhanced neural responses in the dmPFC, a brain region that is implicated in mentalizing. Our findings suggest that mentalizing processes are recruited by the presence of others to attenuate the detection of norm violations and to decrease the level of value calculation of social norms. This idea is consistent with previous observations about the modulations of the mentalizing network on punishment-related brain regions [Baumgartner et al., 2012; Güroğlu et al., 2010; Halko et al., 2009].

We tested this hypothesis by assessing the effective connectivity between punishment- and mentalizing-related networks. Our findings indicated a consistent connectivity pattern that was induced by the presence of others, showing that dmPFC acted as the driver of the network. In light of previous findings, dmPFC is implicated in overtly thinking about the internal mental states of others (i.e., mentalizing) during intergroup interactions [Baumgartner et al., 2012, 2013]. These mentalizing processes play a critical role in diffusion of responsibility, such that individuals often assume that other bystanders must be intervening, and as a result, consider their own altruistic behaviors redundant [Darley and Latané, 1968].

With direct connectivity between dmPFC and target regions (i.e., AI, vmPFC, and precuneus), mentalizing processes implemented by dmPFC modulated neural responses of AI, vmPFC and precuneus, all of which have been previously implicated in punishment behaviors [Baumgartner et al., 2011; Feng et al., 2015; Gu et al., 2015; Xiang et al., 2013]. The AI is a key region that mediates altruistic punishments [Grecucci et al., 2013; Harlé et al., 2012] by encoding deviations from social norms [Civai et al., 2010, 2012]. The current study confirms the context-dependent responses of AI to norm violations [Harlé et al., 2012]. The activity of vmPFC is associated with computing values of fairness, such that the lower BOLD responses of vmPFC to unfair splits than to fair splits might reflect more negative values of norm violations [Aoki et al., 2014; Moretti et al., 2009; Tabibnia et al., 2008; Tricomi et al., 2010; Xiang et al., 2013]. Likewise, precuneus often shows stronger responses to fair splits than to unfair splits [Feng et al., 2015; Xiang et al., 2013]. The involvement of the precuneus in reward processing has been identified in previous studies [e.g., Ballard and Knutson, 2009; Barman et al., 2015; Levy and Glimcher, 2011; McClure et al., 2007], leading to the conjecture that it might participate in encoding positive values of social norms [Feng et al., 2015; White et al., 2014]. Taken together, the current findings suggest that responsibility diffusion in the presence of others diminishes encoding of norm violations and decreases levels of value calculation for perceived splits, which are accomplished by the modulations of the mentalizing-related network.

Several limitations related to our study should be noted. First, behavioral studies investigating diffusion of responsibility often employ one-shot and between-subjects designs, whereas we implemented a multiple-round and within-subjects design to increase statistical power. Moreover, in our study altruistic punishment was clearly prompted by the dictators' splits and participants were explicitly told that they could not see others' decisions and vice versa. The aim of these manipulations was to control for confounding effects of ambiguity and audience inhibition (i.e., running the risk of embarrassment in front of others) that could also account for decreased altruistic behaviors in the presence of others [Latané and Nida, 1981].

Second, it is likely that norm violators could have been punished more severely in the group context than in the alone context, if there had been five real punishers in the group context. However, previous literature investigating diffusion of responsibility usually focuses on how the presence of others decreases altruistic behaviors from *an individual* [Fischer et al., 2011; Wiesenthal et al., 1983], but the overall likelihood/amount of help received by a victim does not necessarily decrease with the increased

availability of helpers [Latané and Nida, 1981]. Moreover, it is plausible that participants reduced punishment to avoid severe monetary loss of transgressors in the group context. Although future studies are needed to investigate this assumption, the current study sufficiently demonstrated that diffusion of responsibility serving as a mediator of context-dependent altruistic punishment plays a causal role in the adjustment of altruistic punishment of norm violations in the presence of others.

Third, negative parameter estimates were observed for several brain regions in our study. In particular, parameter estimates for vmPFC and precuneus were negative in all conditions, indicating a "deactivation" instead of an "activation." Previous studies on decision-making have observed both vmPFC activations [Aoki et al., 2014; Tabibnia et al., 2008; Tricomi et al., 2010] and deactivations [Rao et al., 2014; Sakaiya et al., 2013; Xiang et al., 2013; Zaki and Mitchell, 2011]. Notably, deactivations of vmPFC do not confound the interpretation of its functions. For example, a recent study using a similar economic game paradigm showed an association between vmPFC deactivations and subjective values of perceived splits [Xiang et al., 2013].

Finally, we applied GCM, a data-driven analysis method, to determine the underlying effectively connected brain network, due to the advantage of this method in assessing directional influences among simultaneously recorded BOLD time series in the absence of a priori model of brain connectivity. However, future replications are needed that employ more traditional hypothesis-driven effective connectivity analysis methods (e.g., DCM).

In summary, diffusion of responsibility in the presence of others is an important phenomenon that influences people's decisions on a wide range of social behaviors. Our findings identify the neural basis underlying the modulation of diffusion of responsibility on altruistic punishment and highlight the role of the mentalizing network in this phenomenon. The presence of putative other third-party decision-makers led to a diffusion of responsibility and hence, to a reduction in altruistic punishment to social norm violations. Underlying this effect was a network of interconnected brain regions, with dmPFC acting as the driver of the network and modulating AI, vmPFC and precuneus as the target regions. Our findings have significant implications for various disciplines studying the diffusion of responsibility, including influential phenomena historically observed in social psychology, such as the bystander effect and the "tragedy of the commons".

ACKNOWLEDGMENTS

The authors thank Tengxiang Tian and Xue Feng for their assistance with data collection and Dr. Bobby Azarian for improving language.

REFERENCES

Abler B, Roebroeck A, Goebel R, Höse A, Schönfeldt-Lecuona C, Hole G, Walter H (2006): Investigating directed influences between activated brain areas in a motor-response task using fMRI. Magn Reson Imaging 24:181–185.

Aoki R, Matsumoto M, Yomogida Y, Izuma K, Murayama K, Sugiura A, Camerer CF, Adolphs R, Matsumoto K (2014): Social equality in the number of choice options is represented in the ventromedial prefrontal cortex. J Neurosci 34:6413–6421.

Ballard K, Knutson B (2009): Dissociable neural representations of future reward magnitude and delay during temporal discounting. Neuroimage 45:143–150.

Barman A, Richter S, Soch J, Deibele A, Richter A, Assmann A, Wüstenberg T, Walter H, Seidenbecher CI, Schott BH (2015): Gender-specific modulation of neural mechanisms underlying social reward processing by Autism Quotient. Soc Cognit Affect Neurosci 10:1537–1547.

Baumgartner T, Götte L, Gügler R, Fehr E (2012): The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. Hum Brain Mapp 33:1452–1469.

Baumgartner T, Knoch D, Hotz P, Eisenegger C, Fehr E (2011): Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nat Neurosci 14:1468–1474.

Baumgartner T, Schiller B, Hill C, Knoch D (2013): Impartiality in humans is predicted by brain structure of dorsomedial prefrontal cortex. Neuroimage 81:317–324.

Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300.

Bernhard H, Fischbacher U, Fehr E (2006): Parochial altruism in humans. Nature 442:912–915.

Brainard DH (1997): The psychophysics toolbox. Spatial Vis 10: 433–436.

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

Chang LJ, Sanfey AG (2013): Great expectations: Neural computations underlying the use of social norms in decision-making. Soc Cognit Affect Neurosci 8:277–284.

Civai C, Corradi-Dell'Acqua C, Gamer M, Rumiati RI (2010): Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. Cognition 114:89–95.

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

Civai C, Miniussi C, Rumiati RI (2014): Medial prefrontal cortex reacts to unfairness if this damages the self: A tDCS study. Social Cognit Affect Neurosci 10:1054–1060.

Civai C, Rumiati RI, Rustichini A (2013): More equal than others: Equity norms as an integration of cognitive heuristics and contextual cues in bargaining games. Acta Psychol 144: 12–18.

Corradi-Dell'Acqua C, Civai C, Rumiati RI, Fink GR (2013): Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: An fMRI study. Soc Cognit Affect Neurosci 8:424–431.

Darley JM, Latané B (1968): Bystander intervention in emergencies: diffusion of responsibility. J Person Soc Psychol 8:377.

Davey CE, Grayden DB, Gavrilescu M, Egan GF, Johnston LA (2013): The equivalence of linear gaussian connectivity techniques. Hum Brain Mapp 34:1999–2014.

David O, Guillemain I, Saillet S, Reyt S, Deransart C, Segebarth C, Depaulis A (2008): Identifying neural drivers with functional MRI: An electrophysiological validation. PLoS Biol 6:e315.

Deshpande G, Hu X, Stilla R, Sathian K (2008): Effective connectivity during haptic perception: A study using Granger causality analysis of functional magnetic resonance imaging data. Neuroimage 40:1807–1814.

Deshpande G, LaConte S, James GA, Peltier S, Hu X (2009): Multivariate Granger causality analysis of fMRI data. Hum Brain Mapp 30:1361–1373.

Deshpande G, Libero LE, Sreenivasan KR, Deshpande HD, Kana RK (2013): Identification of neural connectivity signatures of autism using machine learning. Front Hum Neurosci 7:1–15.

Deshpande G, Sathian K, Hu X (2010): Effect of hemodynamic variability on Granger causality analysis of fMRI. Neuroimage 52:884–896.

Engel C (2011): Dictator games: A meta study. Exp Econ 14:583–610.

Fehr E, Camerer CF (2007): Social neuroeconomics: The neural circuitry of social preferences. Trends Cognit Sci 11:419–427.

Fehr E, Fischbacher U (2004a): Social norms and human cooperation. Trends Cognit Sci 8:185–190.

Fehr E, Fischbacher U (2004b): Third-party punishment and social norms. Evol Hum Behav 25:63–87.

Fehr E, Gächter S (2000): Cooperation and punishment in public goods experiments. Am Econ Rev 90:980–994.

Feng C, Luo YJ, Krueger F (2015): Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. Hum Brain Mapp 36:591–602.

Fischer P, Krueger JI, Greitemeyer T, Vogrincic C, Kastenmüller A, Frey D, Heene M, Wicher M, Kainbacher M (2011): The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. Psychol Bull 137:517–537.

Freeman S, Walker MR, Borden R, Latane B (1975): Diffusion of responsibility and restaurant tipping: Cheaper by the bunch. Person Soc Psychol Bull 1:584–587.

Friston K, Moran R, Seth AK (2013): Analysing connectivity with Granger causality and dynamic causal modelling. Curr Opin Neurobiol 23:172–178.

Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. Neuroimage 19:1273–1302.

Granger CW (1969): Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438.

Grant MM, White D, Hadley J, Hutcheson N, Shelton R, Sreenivasan K, Deshpande G (2014): Early life trauma and directional brain connectivity within major depression. Hum Brain Mapp 35:4815–4826.

Grant MM, Wood K, Sreenivasan K, Wheelock M, White D, Thomas J, Knight DC, Deshpande G (2015): Influence of early life stress on intra-and extra-amygdaloid causal connectivity. Neuropsychopharmacology 40:1782–1793.

Grecucci A, Giorgetta C, van't Wout M, Bonini N, Sanfey AG (2013): Reappraising the ultimatum: an fMRI study of emotion regulation and decision making. Cereb Cortex 23:399–410.

Gu X, Wang X, Hula A, Wang S, Xu S, Lohrenz TM, Knight RT, Gao Z, Dayan P, Montague PR (2015): Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: Computational and lesion evidence in humans. J Neurosci 35:467–473.

Guerin B (2011): Diffusion of Responsibility. The Encyclopedia of Peace Psychology, D.J. Christie, Ed.: 336–340. Wiley-Blackwell Publishing Ltd.

Güroğlu B, van den Bos W, Rombouts SA, Crone EA (2010): Unfair? It depends: Neural correlates of fairness in social context. Soc Cognit Affect Neurosci 5:414–423.

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

Handwerker DA, Ollinger JM, D'Esposito M (2004): Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21:1639–1651.

Harlé KM, Chang LJ, van't Wout M, Sanfey AG (2012): The neural mechanisms of affect infusion in social economic decision-making: A mediating role of the anterior insula. Neuroimage 61:32–40.

Harlé KM, Sanfey AG (2012): Social economic decision-making across the lifespan: An fMRI investigation. Neuropsychologia 50:1416–1424.

Havlicek M, Friston KJ, Jan J, Brazdil M, Calhoun VD (2011): Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. Neuroimage 56:2109–2128.

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

Hu J, Blue P, Yu H, Gong X, Xiang Y, Jiang C, Zhou X (2015): Social status modulates the neural response to unfairness. Soc Cognit Affect Neurosci. [Epub ahead of print]

Hu J, Cao Y, Blue PR, Zhou X (2014): Low social status decreases the neural salience of unfairness. Front Behav Neurosci 8: 1–12.

Hutcheson NL, Sreenivasan KR, Deshpande G, Reid MA, Hadley J, White DM, Ver Hoef L, Lahti AC (2015): Effective connectivity during episodic memory retrieval in schizophrenia participants before and after antipsychotic medication. Hum Brain Mapp 36:1442–1457.

Judd CM, Kenny DA, McClelland GH (2001): Estimating and testing mediation and moderation in within-subject designs. Psychol Methods 6:115.

Kahneman D, Knetsch JL, Thaler RH (1986): Fairness and the assumptions of economics. J Business 59:S285–S300.

Kapogiannis D, Deshpande G, Krueger F, Thornburg MP, Grafman JH (2014): Brain networks shaping religious belief. Brain Connectivity 4:70–79.

Katwal SB, Gore JC, Gatenby JC, Rogers BP (2013): Measuring relative timings of brain activities using fMRI. Neuroimage 66: 436–448.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009): Circular analysis in systems neuroscience: The dangers of double dipping. Nat Neurosci 12:535–540.

Krueger F, Landgraf S, van der Meer E, Deshpande G, Hu X (2011): Effective connectivity of the multiplication network: A functional MRI and multivariate Granger Causality Mapping study. Hum Brain Mapp 32:1419–1431.

Lacey S, Stilla R, Sreenivasan K, Deshpande G, Sathian K (2014): Spatial imagery in haptic shape perception. Neuropsychologia 60:144–158.

Latané B, Nida S (1981): Ten years of research on group size and helping. Psychol Bull 89:308–324.

Levy DJ, Glimcher PW (2011): Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. J Neurosci 31:14693–14707.

Lieberman MD, Cunningham WA (2009): Type I and Type II error concerns in fMRI research: Re-balancing the scale. Soc Cognit Affect Neurosci 4:423–428.

Lohmann G, Erfurth K, Müller K, Turner R (2012): Critical comments on dynamic causal modelling. Neuroimage 59:2322–2329.

McClure SM, Ericson KM, Laibson DI, Loewenstein G, Cohen JD (2007): Time discounting for primary rewards. J Neurosci 27: 5796–5804.

Menon V, Uddin LQ (2010): Saliency, switching, attention and control: A network model of insula function. Brain Struct Funct 214:655–667.

Moretti L, Dragone D, Di Pellegrino G (2009): Reward and social valuation deficits following ventromedial prefrontal damage. J Cognit Neurosci 21:128–140.

Mumford JA, Poldrack RA (2014): Adjusting Mean Activation for Reaction Time Effects in BOLD fMRI. OHBM Poster in Hamburg, Germany. Available at: https://ww4.aievolution.com/hbm1401/files/content/abstracts/43589/2053_Mumford.pdf last accessed date: 11/21/2015.

Mynatt C, Sherman SJ (1975): Responsibility attribution in groups and individuals: A direct test of the diffusion of responsibility hypothesis. J Person Soc Psychol 32:1111.

Pelli DG (1997): The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vis 10:437–442.

Poldrack RA (2007): Region of interest analysis for fMRI. Soc Cognit Affect Neurosci 2:67–70.

Poldrack RA, Mumford JA, Nichols TE (2011): Handbook of Functional MRI Data Analysis. Cambridge: Cambridge University Press.

Rao LL, Zhou Y, Liang ZY, Rao H, Zheng R, Sun Y, Tan C, Xiao Y, Tian ZQ, Chen XP (2014): Decreasing ventromedial prefrontal cortex deactivation in risky decision making after simulated microgravity: Effects of −6° head-down tilt bed rest. Front Behav Neurosci 8:1–9.

Roebroeck A, Formisano E, Goebel R (2005): Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25:230–242.

Rosenblatt A, Greenberg J, Solomon S, Pyszczynski T, Lyon D (1989): Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. J Person Soc Psychol 57:681.

Sakaiya S, Shiraito Y, Kato J, Ide H, Okada K, Takano K, Kansaku K (2013): Neural correlate of human reciprocity in social interactions. Front Neurosci 7.

Sanfey AG, Chang LJ (2008): Multiple systems in decision making. Ann NY Acad Sci 1128:53–62.

Sanfey AG, Loewenstein G, McClure SM, Cohen JD (2006): Neuroeconomics: Cross-currents in research on decision-making. Trends Cognit Sci 10:108–116.

Sathian K, Deshpande G, Stilla R (2013): Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. J Neurosci 33:5387–5398.

Sebastian CL, Fontaine NM, Bird G, Blakemore SJ, De Brito SA, McCrory EJ, Viding E (2012): Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. Soc Cognit Affect Neurosci 7:53–63.

Sreenivasan KR, Havlicek M, Deshpande G (2015): Non-Parametric Hemodynamic Deconvolution of fMRI using Homomorphic Filtering. IEEE Trans Med Imaging 34:1155–1163.

Stilla R, Deshpande G, LaConte S, Hu X, Sathian K (2007): Posteromedial parietal cortical activity and inputs predict tactile spatial acuity. J Neurosci 27:11091–11102.

Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, Kirsch P (2011): Beyond revenge: neural and genetic bases of altruistic punishment. Neuroimage 54:671–680.

Tabibnia G, Satpute AB, Lieberman MD (2008): The sunny side of fairness preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). Psychol Sci 19:339–347.

Tricomi E, Rangel A, Camerer CF, O'Doherty JP (2010): Neural evidence for inequality-averse social preferences. Nature 463: 1089–1091.

Uddin LQ, Supekar K, Lynch CJ, Cheng KM, Odriozola P, Barth ME, Phillips J, Feinstein C, Abrams DA, Menon V (2014): Brain state differentiation and behavioral inflexibility in autism. Cereb Cortex 25:4740–4747.

van Bommel M, van Prooijen JW, Elffers H, Van Lange PA (2012): Be aware to care: Public self-awareness leads to a reversal of the bystander effect. Journal of Experimental Social Psychology, 48:926–930.

Vul E, Harris C, Winkielman P, Pashler H (2009): Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect Psychol Sci 4:274–290.

Vul E, Kanwisher N (2010): Begging the question: The non-independence error in fMRI data analysis. In: Hanson, S.B.M. (Ed.), Foundational Issues for human brain mapping. MIT Press, Cambridge, MA

Wegner DM, Schaefer D (1978): The concentration of responsibility: An objective self-awareness analysis of group size effects in helping situations. J Person Soc Psychol 36:147.

Wheelock M, Sreenivasan K, Wood K, Ver Hoef L, Deshpande G, Knight D (2014): Threat-related learning relies on distinct dorsal prefrontal cortex network connectivity. Neuroimage 102: 904–912.

White SF, Brislin SJ, Sinclair S, Blair JR (2014): Punishing unfairness: rewarding or the organization of a reactively aggressive response? Hum Brain Mapp 35:2137–2147.

Wiesenthal DL, Austrom D, Silverman I (1983): Diffusion of responsibility in charitable donations. Basic Appl Soc Psychol 4:17–27.

Wright ND, Symmonds M, Fleming SM, Dolan RJ (2011): Neural segregation of objective and contextual aspects of fairness. J Neurosci 31:5244–5252.

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. Soc Cognit Affect Neurosci 9:1862–1871.

Xia M, Wang J, He Y (2013): BrainNet Viewer: A network visualization tool for human brain connectomics. PLoS One 8:e68910.

Xiang T, Lohrenz T, Montague PR (2013): Computational substrates of norms and their violations during social exchange. J Neurosci 33:1099–1108.

Yu R, Calder AJ, Mobbs D (2014): Overlapping and distinct representations of advantageous and disadvantageous inequality. Hum Brain Mapp 35:3290–3301.

Yu R, Hu P, Zhang P (2015): Social distance and anonymity modulate fairness consideration: An ERP study. Sci Rep 5:1–12.

Zaki J, Mitchell JP (2011): Equitable decision making is associated with neural markers of intrinsic value. Proc Natl Acad Sci USA 108:19761–19766.