Neural Substrates of Context- and Person-Dependent Altruistic Punishment

Lili Wang,[1] Xiaping Lu,[2] Ruolei Gu,[3] Ruida Zhu,[2] Rui Xu,[4] Lucas S. Broster,[5] and Chunliang Feng [2,6]*

[1]*School of Educational Science, Huaiyin Normal University, Huaian, China*
[2]*State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China*
[3]*Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*
[4]*Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China*
[5]*Department of Psychiatry, University of California at San Francisco, San Francisco, California*
[6]*College of Information Science and Technology, Beijing Normal University, Beijing, China*

Abstract: Human altruistic behaviors are heterogeneous across both contexts and people, whereas the neural signatures underlying the heterogeneity remain to be elucidated. To address this issue, we examined the neural signatures underlying the context- and person-dependent altruistic punishment, conjoining event-related fMRI with both task-based and resting-state functional connectivity (RSFC). Acting as an impartial third party, participants decided how to punish norm violators either alone or in the presence of putative others. We found that the presence of others decreased altruistic punishment due to diffusion of responsibility. Those behavioral effects paralleled altered neural responses in the dorsal anterior cingulate cortex (dACC) and putamen. Further, we identified modulation of responsibility diffusion on task-based functional connectivity of dACC with the brain regions implicated in reward processing (i.e., posterior cingulate cortex and amygdala/orbital frontal cortex). Finally, the RSFC results revealed that (i) increased intrinsic connectivity strengths of the putamen with temporoparietal junction and dorsolateral PFC were associated with attenuated responsibility diffusion in altruistic punishment and (ii) increased putamen-dorsomedial PFC connectivity strengths were associated with reduced responsibility diffusion in self-reported responsibility. Taken together, our findings elucidate the context- and person-dependent altruistic behaviors as well as associated neural substrates and thus provide a potential neurocognitive mechanism of

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

Conflict of Interest: The authors are unaware of any conflicts of interest, financial or otherwise.

Contract grant sponsor: National Natural Science Foundation of China; Contract grant numbers: 31300869, 31571124, 81503480; Contract grant sponsor: Nature Science Foundation of JiangSu Province; Contract grant number: BK20130415; Contract grant sponsor: National Postdoctoral Program for Innovative Talents; Contract grant number: BX201600019; Contract grant sponsor: Beijing National Science Foundation; Contract grant number: 7154227; Contract grant sponsor: National Institutes of Health; Contract grant number: R25 MH060482; Contract grant sponsor: Qing Lan Project

*Correspondence to: Chunliang Feng, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China. E-mail: chunliang.feng@bnu.edu.cn

Received for publication 21 June 2017; Revised 13 July 2017; Accepted 17 July 2017.

DOI: 10.1002/hbm.23747

Published online 26 July 2017 in Wiley Online Library (wileyonlinelibrary.com).

heterogeneous human altruistic behaviors. *Hum Brain Mapp 38:5535–5550, 2017.*

Key words: altruistic punishment; responsibility diffusion; functional magnetic resonance imaging; psychophysiological interaction; resting-state functional connectivity

INTRODUCTION

Altruism constitutes one of the defining features of human societies [Fehr and Fischbacher, 2003], such that humans exhibit preferences in sharing and cooperation that are at odds with maximizing personal payout [Engel, 2011; Fehr et al., 2008]. Further, people often sacrifice some of their own resources to punish violators of social norms, a behavioral tendency termed as altruistic punishment [Fehr and Fischbacher, 2004b; Fehr and Gächter, 1999, 2000, 2002]. People punish norm violators not only in the situations where they themselves are victims (i.e., second-party punishment) but also when they are unaffected bystanders (i.e., third-party punishment, TPP) [Baumgartner et al., 2012; Fehr and Gächter, 1999; Feng et al., 2016]. These findings together corroborate the notion that human beings are altruistic. However, human altruistic behaviors are not ubiquitous, but are heterogeneous across both social contexts and individuals [Civai et al., 2012; Feng et al., 2015a,; Fischbacher et al., 2001; Haruno et al., 2014].

The presence of others is a critical social context that substantially modulates people's altruistic behaviors [Fischer et al., 2011]. The presence of others often leads to decreased amounts or likelihood of help [Latané and Darley, 1968; van Bommel et al., 2012], donation [Wegner and Schaefer, 1978; Wiesenthal et al., 1983] and altruistic punishment [Chekroun and Brauer, 2002; Feng et al., 2016], due to the diffusion of responsibility. Specifically, diffusion in responsibility is likely to constrain altruistic behaviors when (i) other people are present and are potentially available to intervene and (ii) the behaviors of others cannot be closely observed [Latané and Darley, 1968]. Additionally, people's willingness to engage in altruistic behavior is modulated by a variety of personal attributes, including empathic ability [Eisenberg and Miller, 1987; Masten et al., 2011; Mathur et al., 2010; Morelli et al., 2014; Tusche et al., 2016], narcissistic traits [Böckler et al., 2017], agreeableness [Martin-Raugh et al., 2016; Oda et al., 2014], and genotypes [Crişan et al., 2009; McDermott et al., 2009]. For instance, the more bystanders empathize with the distress of others, the more likely they engage in altruistic behaviors [Batson et al., 1986; Hein et al., 2010; Rameson et al., 2012]. Although the context- and person-dependent nature of human altruism has long been recognized, the underlying neural substrates remain understudied.

In this study, we aimed to address this issue combining a TPP task with event-related fMRI and functional connectivity. The TPP task has frequently been used to probe altruistic punishment and neural responses to norm violations that do not directly affect oneself [Baumgartner et al., 2012; Fehr and Fischbacher, 2004b]. Previous studies utilizing the altruistic punishment task have revealed the consistent involvement of a salience network, anchored in the dorsal anterior cingulate cortex (dACC) and anterior insula (AI) [Feng et al., 2015b; Gabay et al., 2014]. These brain regions have been implicated in detecting norm-violating behaviors [Feng et al., 2017; Strobel et al., 2011] as well as predicting decisions to punish violators [Gabay et al., 2014]. Further, those punishment-related neural responses are modulated according to social contexts and people [Baumgartner et al., 2012; Haruno and Frith, 2010; Haruno et al., 2014; Wright et al., 2011]. The influence of social contexts or personal attributes on altruistic punishment is associated with neuropsychological processes implemented in prefrontal areas, including the dorsomedial prefontal cortex (dmPFC) and dorsolateral PFC (dlPFC), which are respectively implicated in mentalizing and integrating different sources of information to optimize decision-making [Baumgartner et al., 2012; Buckholtz et al., 2008; Feng et al., 2016; Halko et al., 2009].

Furthermore, this study took a novel approach to explore individual differences in altruistic punishment by examining the potential contribution of the brain's intrinsic functional architecture—measured as resting-state functional connectivity (RSFC)—on human punishment behavior. RSFC has emerged as a system-level approach for delineating brain function and neural basis of cognitive ability and personality traits since the seminal work of Biswal et al. [1995], who first demonstrated coherent low-frequent fluctuations of the somatomotor system in the resting-state BOLD signal. Reliable patterns of resting-state coherence (i.e., RSFC) that mirror those engaged by goal-directed tasks have since been identified within many well-characterized networks [Raichle, 2011, 2015]. Therefore, spontaneous fluctuations of BOLD signal in functionally related brain regions are not random, but are intrinsically organized, which offers a potential predictor of relevant behaviors [Harmelech and Malach, 2013]. Indeed, individual differences in the strength of RSFC within a variety of brain networks, including somatomotor [Taubert et al., 2011], perception [Lewis et al., 2009], attention [Mennes et al., 2010], and knowledge systems [Wang et al., 2013; Wei et al., 2012], contribute significantly to individual variations in associated behavioral performance. These findings bolster an intrinsic perspective of brain function, which argues that intrinsic neural activity

"instantiates the maintenance of information for interpreting, responding to, and even predicting environmental demands" [Raichle, 2006]. Whether the intrinsic RSFC-behavior link persists for more complex social interactions such as altruistic punishment remains unknown [Birn, 2007], and our work represents an initial effort to tackle this question.

In light of previous findings, we hypothesized that the presence of others would result in decreases in altruistic punishment due to a diffusion of responsibility. The responsibility diffusion in altruistic punishment would be associated with changes in neural activity in the salience network consisting of the dACC and AI. Further, we hypothesized that those context-dependent behavioral and neural responses to norm violations would be related to modulation of mentalizing and information-integration processes (e.g., dmPFC, dlPFC) during the TPP task. Finally, we expected that the intrinsic functional connectivity between the salience network and mentalizing/information-integration network would underlie individual differences in the responsibility diffusion of altruistic punishment. Taken together, our study examined both context- and person-dependent altruistic punishment and their neural signatures.

MATERIALS AND METHODS

Subjects

Twenty-six students (15 females; mean age $\pm$ s.d. $= 20.92 \pm 2.04$ years) participated in the study for monetary compensation. All participants were right-handed, had normal or corrected-to-normal vision, and had 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).

Experimental Procedure and Task

Participants underwent three sessions for this study. First, groups of 3–4 participants were invited to the laboratory for a "screening session" a week prior to the fMRI scanning. Participants were informed that the upcoming fMRI session would be conducted in groups of three people, with one participant in the MRI-scanning room and two other participants in a room equipped with two computers near the MRI-scanning room.

Second, participants returned the following week for the *fMRI* "scanning session" to play a TPP game. As third-party decision-makers (player C), participants observed how a sum of money (12 MUs) was allocated between several pairs of other players (A and B). In particular, 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 another 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 [Kahneman et al., 1986]. On each round, participants were given 6 MUs and had a chance to reduce A's payoff as a punishment by altruistically spending their own money: each MU spent reduced 2 MUs from A's payoff [Bernhard et al., 2006; Fehr and Fischbacher, 2004b]. It is noteworthy that terms such as "fairness," "punish," or "sanction" were never employed in the instructions. Instead, we followed a conventional approach, telling participants that they had the chance to assign "deduction points" to the proposers [Fehr and Fischbacher, 2004b].

Participants made their decisions under both "alone" and "group" contexts (Fig. 1a,b). In the alone context, only decisions of one player ("A") would be presented to participants, and participants decided whether and by how much to reduce A's payoff. For instance, if the participant decided to spend 1 MU, then 2MUs would be reduced from player A's payoff. In the group context, there were three pairs of players (organized as three pairs of players "A and B") such that player A in each pair had made the same allocation. The decisions of these "A" players were presented to participants and two other putative C players, who were understood to be performing the same task outside the scanner. In this context, the total amounts of MUs spent by the three C players would be split to reduce each A player's payoff equally. For instance, if three C players spent 3 MUs in total, then 2 MUs would be reduced from each A player's payoff. Notably, three pairs of players A and B (rather than one pair) were introduced in the group context to control for efficiency of punishment between alone and group contexts. As such, decreases in altruistic punishment in the group context could not be attributed to the potential confound that the cumulative punishment from all three C players might be perceived as too severe. However, it is noteworthy that the main findings of this study were replicated by a control experiment in which there was only one "A and B" player pair in the group context (for details, please refer to the Supporting Information text and Supporting Information Fig. 12).

Prior to fMRI scanning, 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. While participants were prepared for the fMRI session by one experimenter, another experimenter showed up to announce that the other two volunteers (i.e., two other putative C players) participating in the behavioral part of the experiment were ready to begin the experiment. Inside the MRI scanner, participants saw instructions asking all (three) players to press a button to begin the experiment, which further explained that the experiment could only begin after all players had pressed their buttons. In reality,

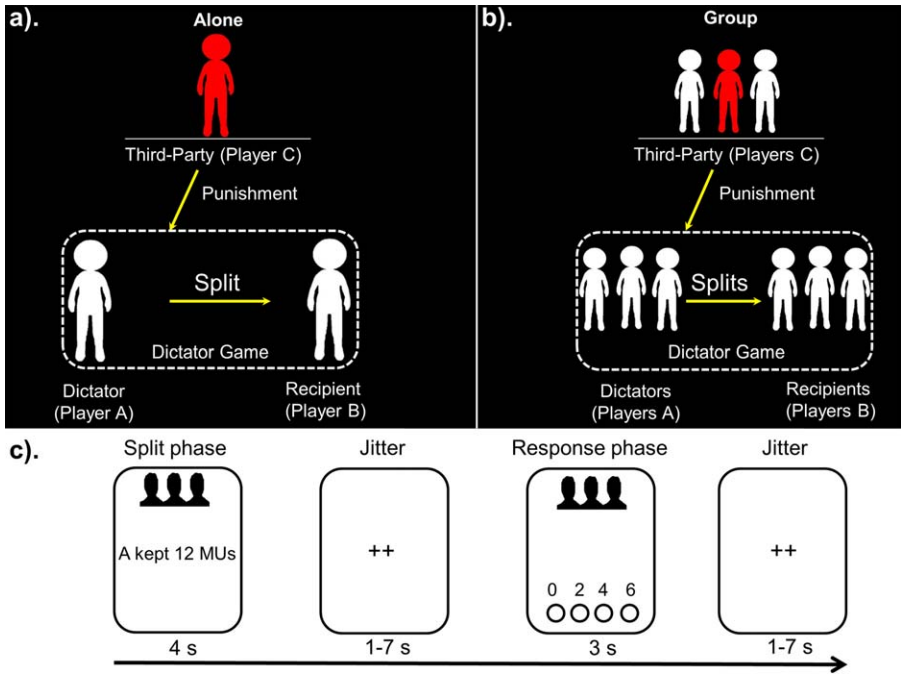

Figure 1.

Task design and experimental procedure. **(a, b)** TPP task in the alone (a) and group (b) contexts. In the alone context, participants acted as a third-party and independently decided how to punish a dictator. In the group context, participants and two other third-party decision-makers together decided how to punish three dictators. **(c)** Experimental procedure. On each round, split (fair vs. unfair) and context information (alone vs. group) was presented and followed by a jitter. Then, participants had to indicate how much money (i.e., punishment points) they were willing to spend to reduce the dictator or dictators' allocation using one of the four possible choices (monetary units: 0, 2, 4, 6). Finally, an optimized jitter was presented. [Color figure can be viewed at wileyonlinelibrary.com]

the software program triggered the other two button responses automatically. Overall, the aim of these procedures was to increase realism so that participants believed that they were playing with two other people.

On each round of TPP game, context information (alone context or group context) and A's allocation were presented constantly for 4 s (Fig. 1c). Afterwards, a jitter was presented (1–7 s) and followed by the response phase (3 s), during which time participants had to indicate how many MUs they were willing to spend to reduce A's payoff using one of the four possible choices (a hollow circle was under each choice): 0, 2, 4, or 6 MUs. Participants made their decisions through a response box, with associations between buttons and decisions being counterbalanced across subjects. Once their response was registered, feedback indicating their decision with filled circle under the corresponding choice was displayed onscreen until the end of the decision period. Each round ended with a second jitter (1–7 s). Stimulus presentation and behavioral data collection were implemented by using Psychtoolbox-3 (http://psychtoolbox.org/).

Afterwards, participants completed two 9-minute runs of TPP game (270 scans every 2 s). Each run consisted of 40 rounds: 4 rounds of 12:0 splits, 3 rounds of 11:1, 3 rounds of 10:2 splits, 5 rounds of 7:5, and 5 rounds of 6:6 splits for both alone and group contexts. To mitigate loss of statistical power, splits of 6:6 and 7:5 were clustered as fair splits, whereas splits of 10:2, 11:1, and 12:0 were clustered as unfair splits (for results assessing each split separately, please refer to the Supporting Information text, Fig. S6–S10, and Table S5). 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, the within-subjects factors (2 [Split: fair, unfair] $\times$ 2 [Context: alone, group]) consisted of 20 rounds per condition, including fair splits in the alone context (Alone-Fair), unfair splits in the alone context (Alone-Unfair), fair splits in the group context (Group-Fair), and unfair splits in the group context (Group-Unfair). All conditions were randomly presented on a trial-by-trial basis.

Finally, in the "post-scan session" participants completed a survey. Participants were asked to rate the same splits (fair: 6:6, 7:5; unfair: 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, 9 = absolutely), "To what extent did you feel that A's allocations were fair?" (Fairness: 1 = absolutely unfair, 9 = absolutely fair), "How excited did you feel" (Emotional arousal ratings: 1 = very calm, 9 = very excited), and "How pleased did you feel" (Emotional valence ratings: 1 = very unpleasant, 9 = very pleasant).

To encourage real decisions from participants, it was emphasized that MUs were convertible to monetary payoff, and that participants 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 US dollars) at the end of experiment [Civai et al., 2014; Corradi-Dell'Acqua et al., 2013; Grecucci et al., 2013]. Before leaving the laboratory, participants completed a debriefing questionnaires designed to examine their beliefs about the experimental setup. No participants expressed doubts as to whether (i) two other players were playing with them outside the scanner; and (ii) payoffs of 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-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 × 224 mm^2. Prior to the fMRI scanning of the TPP game, participants completed a 5-min resting-state fMRI scanning. During resting-state fMRI scanning, participants were instructed to close their eyes, keep still, remain awake, and not think about anything systematically. The resting state scanning consisted of 150 contiguous EPI volumes. The fMRI scanning for the TPP game consisted of 540 EPI volumes in total (2 runs, 270 EPI volumes per run). High-resolution anatomical images covering the entire brain were obtained by applying magnetization-prepared rapid acquisition with a gradient-echo sequence: TR/TE = 2530 ms/3.39 ms, flip angle = 7°, number of slices = 144, slices thickness = 1.33 mm, matrix size = 256 × 256, FOV = 256 × 256 mm^2.

Statistical Analysis

Behavioral data

Behavioral data analyses were performed using SPSS 21.0 (IBM, Somers, USA) with a threshold of $P < 0.05$ (two-tailed). To investigate the effects of social context (i.e., diffused responsibility), repeated measure analysis of variances (ANOVAs) on TPP (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 versus fair 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 details, see Supporting Information text).

Task-based fMRI data: Activation analysis

Neuroimaging data analyses were performed with SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Preprocessing of functional data included realignment through rigid-body registration to correct for head motion, slice timing and normalization to MNI space, interpolation of voxel sizes to 2 × 2 × 2 mm, smoothing (8-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 both split-phase (4 s) and response-phase (3 s) in four conditions: fair splits in the alone and group contexts, unfair splits in the alone and group contexts. 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. For the second level, we focused on those parameter estimates from the first level associated with the four conditions of our 2 × 2 design. These images were then fed into a flexible factorial design with a within-subject factor of four levels using a random effects analysis. In particular, the interaction between Split (fair, unfair) and Context (alone, group) was assessed by calculating the contrast of (Alone [Unfair – Fair] > Group [Unfair – Fair]). Likewise, main effects of Split and Context were assessed by calculating the contrasts of (Unfair – Fair) and (Alone – Group), respectively. Further, simple effects of Split were calculated for both alone and group contexts. We focused on the analysis for

the split phase [see also Guo et al., 2014; Halko et al., 2009; Xiang et al., 2013]. Since all information necessary for decision-making was presented in the split phase, it is likely that it elicited not only the encoding of splits, but also decision-related neuropsychological processes. However, we expect that the separation of split and response phases in this study helped to minimize motor-related confounds for the contrast of interest (i.e., the split phase). Results of the response phase have been presented in the Supporting Information material for the sake of completeness (Supporting Information Table S1 and Fig. S1). 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 $P < 0.05$ [Eklund et al., 2016; Woo et al., 2014].

Task-based fMRI data: Connectivity analysis

The activation analysis identified the involvement of dACC and right putamen in the effect of diffusion in responsibility (see also Results section). Therefore, we examined whether the these areas worked together with other brain regions to underlie the responsibility diffusion, with an analysis of Psychophysiological interaction (PPI) [Friston et al., 1997] using the identified areas as regions of interest (ROIs). 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 each ROI as the seeding signals. These seeding signals were then deconvolved with the canonical HRF, resulting in estimates of 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 reconvolved 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 was similar to that of the activation data, except that the beta values used were derived from the PPI regressors. In this study, we focused on connections that exhibited an interaction between Split and Context (i.e., diffusion of responsibility). Multiple comparisons were corrected with the same approach and thresholds used in the activation analysis.

Resting-state fMRI data: Preprocessing

To further determine the relationship between intrinsic organization of functional networks (using dACC and right putamen as the seed regions) and behavioral responsibility diffusion, voxel-wise correlation analyses were conducted between intrinsic connectivity strength of each voxel with seed regions and context-dependent decisions/responsibility ratings in the TPP.

Resting-state neuroimaging data analyses were performed with SPM12. The functional images were corrected for slice-timing and realigned for head movement correction to the mean image. To normalize functional images, participants' structural brain images were first segmented and then all functional images were coregistered to their own structural images. The parameters derived from segmentation were used to normalize each participant's functional images into MNI space (resampling voxel size was $2 \times 2 \times 2$ mm^3). Subsequently, the images were spatially smoothed using a Gaussian filter (FWHM = 6 mm) to decrease spatial noise.

Resting-state fMRI data: Seed-to-voxel connectivity

Implementing a seed-based analysis, the functional connectivity (bivariate correction) between the average BOLD signals from given seed regions (i.e., dACC and right putamen) and all other voxels in the brain was computed using the Functional Connectivity (CONN) toolbox (https://www.nitrc.org/projects/conn). To remove potential confounds, regressors of no interest were added in the first-level GLM, including six head motion parameters (three translations and three rotations along x, y, and z axes), white matter, and cerebrospinal fluid signal. The Pearson's correlation coefficients obtained at each voxel were transformed into Fisher's z values to indicate the degree of connectivity between each seed region and the voxel.

Resting-state fMRI data: Correlational analyses

First-level, subject-specific, connectivity maps for each ROI were then employed in a second-level analysis in which correlation analyses were performed to determine the associations between behavioral responsibility diffusion (decisions and responsibility ratings) and the functional connectivity (z scores) of the dACC and right putamen as the seed with other regions. To correct for multiple comparisons, FWE corrected P values that were < 0.05 at the cluster level (cluster-defining threshold: $P < 0.001$) were considered significant (implemented in SPM12).

Resting-state fMRI data: Pattern regression analysis

The analysis aimed to test whether participants' diffusion of responsibility could be decoded from resting-state connectivity patterns of dACC and putamen [Fernandes et al., 2017]. This complemented the univariate correlational analysis by providing two primary advantages: (i) the multivariate nature of the analysis enabled the detection of subtle and spatially distributed effects and (ii) the analysis allowed for predicting unseen participants,

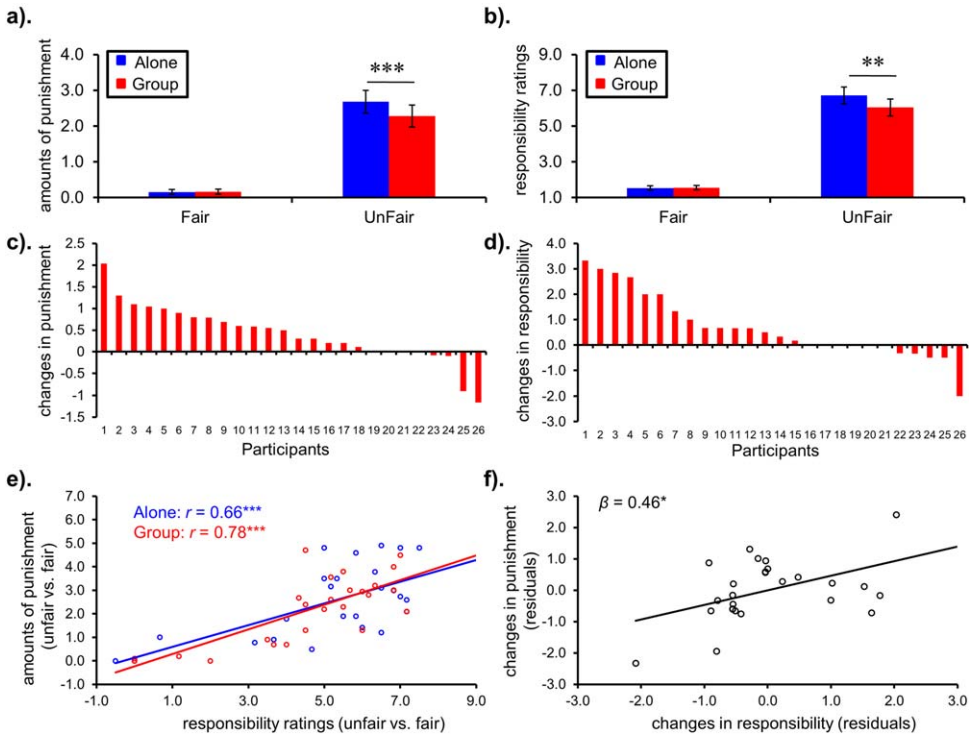

Figure 2.

Diffusion of responsibility in altruistic punishment. **(a, b)** Modulations of the presence of others on altruistic punishment and subjective responsibility. The presence of others attenuated both altruistic punishment behavior and subjective responsibility. **(c, d)** The effects of the presence of others on altruistic punishment and subjective responsibility for each participant. **(e, f)** correlations between subjective responsibility and altruistic punishment. Sense of responsibility predicted altruistic punishment in both alone and group contexts. In addition, responsibility diffusion served as a mediator for the reductions in altruistic punishment in the presence of others. $^{***}P < 0.0005$. $^{**}P < 0.005$. $^{*}P < 0.05$. Error bars indicate standard error. [Color figure can be viewed at wileyonlinelibrary.com]

offering information at the individual rather than at the group level.

The pattern regression analysis was implemented in PRoNTo (http://www.mlnl.cs.ucl.ac.uk/pronto/) and included the following steps [Fernandes et al., 2017; Schrouff et al., 2013]: (1) The resting-state connectivity of dACC and putamen as seeding regions was derived from the seed-to-voxel connectivity analysis. (2) The whole brain was split into 116 anatomical regions according to the *aal* atlas. (3) For each region, a linear kernel or "similarity matrix" was computed according to the connectivity patterns of all voxels within the region. Therefore, a 26 × 26 (i.e., 26 participants) kernel matrix was generated for each region. (4) The kernels from all anatomical regions for both seeding ROIs (i.e., dACC and putamen) were hierarchically combined using Multiple Kernel Learning (MKL) [Schrouff et al., 2014], which aims at simultaneously learning the kernel weights in supervised learning settings. In particular, MKL determines the relative contribution of each region (kernel weights) for the final decision function, as well as the relative contribution of each voxel (voxel weights) within each region. In other words, MKL can be considered a hierarchical model, in which the models corresponding to individual brain regions are assembled to form the final brain model. Given the sparse nature of the MKL implemented in PRoNTo, only a subset of the regions would be selected in the regression analysis. (5) Both leave-one-subject-out cross-validation (LOSOCV) and 10-fold cross-validation procedures were employed to train and examine the performance of the model [Cui et al., 2016; Fernandes et al., 2017]. (6) The performance was assessed by measuring the consistency between the predicted and actual values, using Pearson's correlation coefficient (r). (7) The permutation test was applied to determine the significance of the model's performance. That is, changes in amounts of punishment or responsibility ratings were permutated across the sample ($n = 26$) 1,000 times, and the entire regression procedure was reapplied each time. The P value for the r was calculated by dividing the number of permutations that showed a higher value than the actual value for the real

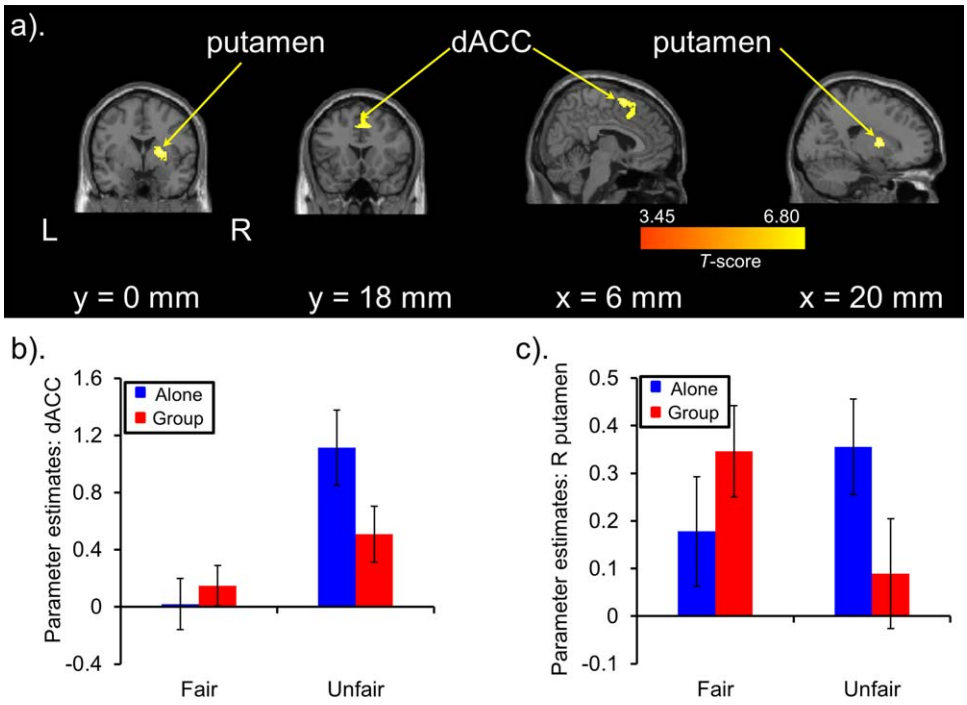

Figure 3.

Influence of the presence of others on neural responses to unfair and fair splits. **(a)** Activation revealed by the interaction of Split and Context at dACC and putamen. **(b, c)** Illustrations of the parameter estimates for the dACC and putamen as a function of Split (fair, unfair) and Context (alone, group). L, left; R, right; dACC, dorsal anterior cingulate cortex. [Color figure can be viewed at wileyonlinelibrary.com]

sample by the total number of permutation (i.e., 1,000). The resulting P values were corrected for multiple comparisons using the approach developed by Benjamini and Hochberg [1995] to control for the false discovery rate [FDR, q(FDR) $<$ 0.05] (https://cn.mathworks.com/matlabcentral/fileexchange/27418-fdr-bh).

To visualize neural patterns for brain regions identified in voxel-wise whole-brain analysis [Poldrack, 2007; Poldrack et al., 2011; Vul and Kanwisher, 2010], parameter estimates of identified brain regions were extracted 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].

RESULTS

Behavioral Results

Decisions and responsibility ratings

The ANOVA on amounts of punishment revealed significant main effects of Split ($F_{1,\ 25} = 65.21$, $P < 0.0005$) and Context ($F_{1,\ 25} = 8.45$, $P < 0.01$), indicating that participants gave stronger punishments in response to unfair splits (norm violations) than to fair splits and stronger punishments in the alone context than in the group context. A significant interaction effect of Split $\times$ Context was observed ($F_{1,\ 25} = 10.23$, $P < 0.005$); post-hoc comparisons revealed that unfair splits were punished more strongly in the alone context than in the group context ($t_{25} = 3.16$, $P < 0.0005$, Fig. 2a). Similar effects of Split, Context, and their interaction were identified for rates of punishment (see also Supporting Information text and Fig. S1 for details).

The ANOVA on responsibility ratings revealed significant main effects of Split ($F_{1,\ 25} = 137.57$, $P < 0.0005$) and Context ($F_{1,\ 25} = 7.81$, $P < 0.05$), 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 $\times$ Context was observed ($F_{1,\ 25} = 7.88$, $P < 0.05$), demonstrating that participants felt more responsible for punishing norm violations (i.e., unfair splits) in the alone context compared to the group context ($t_{25} = 2.93$, $P < 0.01$, Fig. 2b). Although these results demonstrated reliable context-dependent altruistic behaviors that were linked to the diffusion in subjective responsibility, participants differed widely with respect to the degree of diffusion of responsibility (Fig. 2c,d).

Regarding the relationship between subjective responsibility and amounts of punishment, subjective responsibility

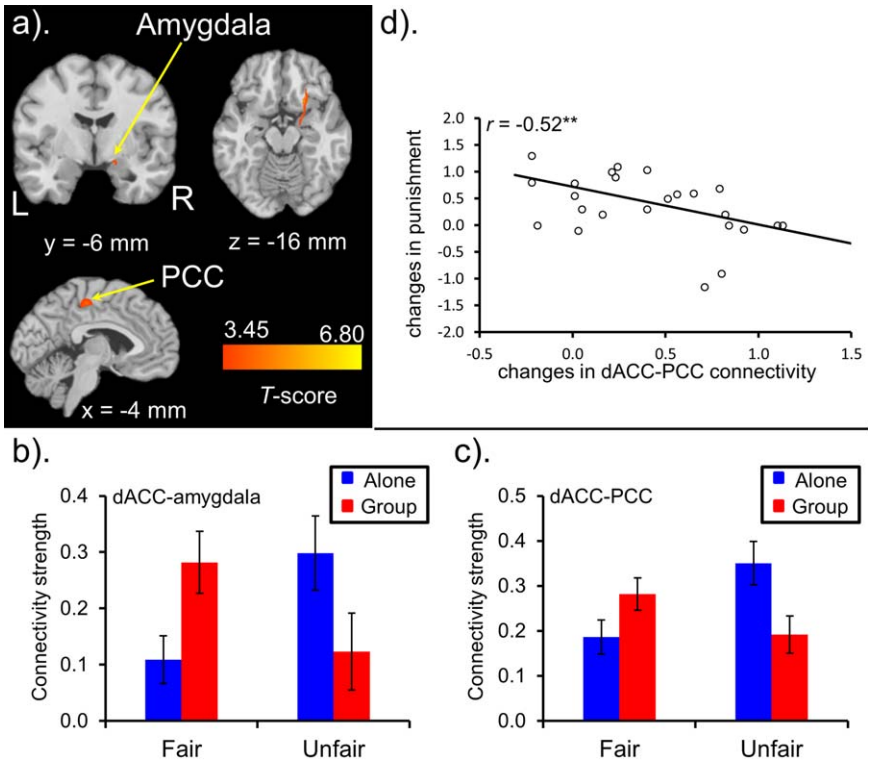

Figure 4.

Influence of the presence of others on functional connectivity in response to norm violations. **(a)** Context-dependent functional connectivity of dACC. The presence of others was associated with attenuated functional couplings of dACC with posterior cingulate cortex and amygdala/orbital frontal cortex. **(b, c)** Illustrations of dACC-amygdala and dACC-PCC connectivity strength as a function of Split (fair, unfair) and Context (alone, group). **(d)** Functional connectivity-behavior correlations. The strength of dACC-PCC connectivity was negatively correlated with responsibility diffusion in altruistic punishment. $^{**}P < 0.01$. L, left; R, right; PCC, posterior cingulate cortex; dACC, dorsal anterior cingulate cortex. [Color figure can be viewed at wileyonlinelibrary.com]

was predictive of punishment to unfair versus fair splits in both alone ($n = 26$, $r = 0.66$, $P < 0.0005$) and group contexts ($n = 26$, $r = 0.78$, $P < 0.0005$; Fig. 2e), suggesting the critical role of sense of responsibility in altruistic punishment. Notably, 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 ($\beta = 0.46$, $t_{25} = 2.51$, $P < 0.05$; Fig. 2f). Taken 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 (for details, see also Supporting Information material).

Control ratings

The ANOVAs on ratings of fairness and emotional feelings (arousal, valence) yielded only a significant main effect of Split (fairness: $F_{1,\ 25} = 960.76$, $P < 0.0005$; emotional arousal: $F_{1,\ 25} = 149.53$, $P < 0.0005$, and emotional valence: $F_{1,\ 25} = 252.97$, $P < 0.0005$), demonstrating that participants' impressions of fairness and feelings of pleasantness were lower for unfair splits than for fair splits, whereas participants felt more emotionally aroused in response to unfair splits than to fair splits. However, there was neither a significant main effect of Context (fairness: $F_{1,\ 25} = 0.001$, $P > 0.05$; emotional arousal: $F_{1,\ 25} = 1.34$, $P > 0.05$, and emotional valence: $F_{1,\ 25} = 0.35$, $P > 0.05$) nor a significant interaction effect of Split × Context (fairness: $F_{1,\ 25} = 0.01$, $P > 0.05$; emotional arousal: $F_{1,\ 25} = 0.19$, $P > 0.05$, and emotional valence: $F_{1,\ 25} = 0.14$, $P > 0.05$; Supporting Information Fig. S1).

Task-Based fMRI Activation Results

The contrast of (Alone [Unfair – Fair] > Group [Unfair – Fair]) revealed changes in BOLD responses in the dACC ($x/y/z = 8/16/44$ mm) and right putamen ($x/y/z = 24/2/$

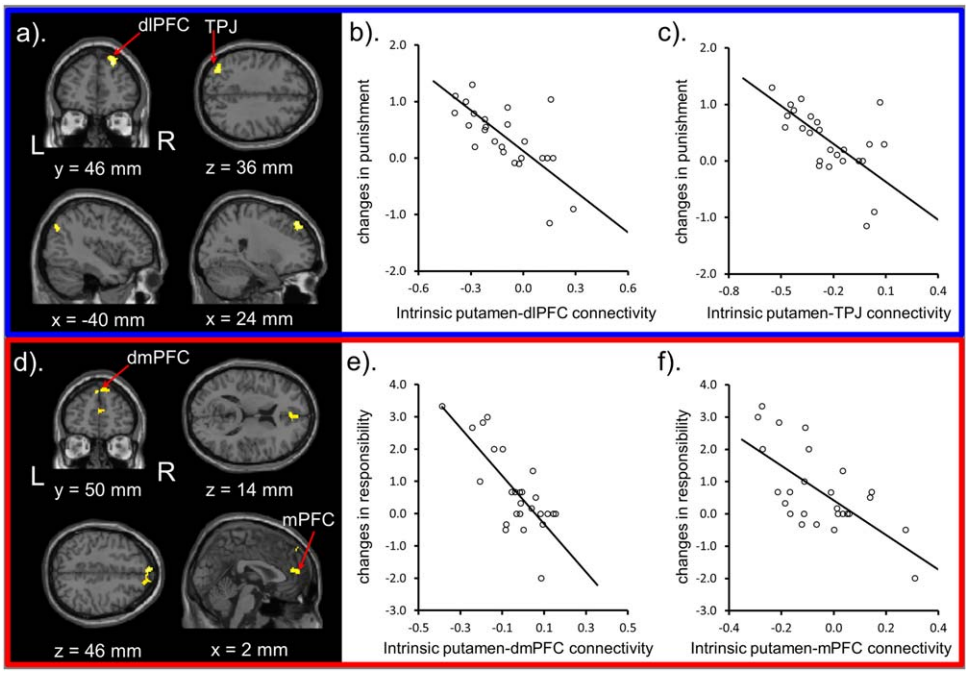

Figure 5.

Brain regions for which functional connectivity strength with putamen was significantly correlated with responsibility diffusion in punishment behavior and responsibility ratings. **(a)** dlPFC and TPJ were implicated as regions with which functional connectivity strength with putamen was significantly correlated with responsibility diffusion in punishment behaviors. **(b, c)** Scatterplots show negative correlations between responsibility diffusion in punishment behavior and functional connectivity strength between putamen and each cluster. Each dot represents data from one participant. **(d)** Regions for which functional connectivity strength with putamen was significantly correlated with responsibility diffusion in responsibility ratings. The correlation analyses revealed mPFC and dmPFC. **(e, f)** Scatterplots show negative correlations between responsibility diffusion in responsibility ratings and functional connectivity strength between putamen and each cluster. Each dot represents data from one participant. L, left; R, right; dlPFC, dorsolateral prefrontal cortex; TPJ, temporoparietal junction; dmPFC, dorsomedial prefrontal cortex; mPFC, medial prefrontal cortex. [Color figure can be viewed at wileyonlinelibrary.com]

8 mm; Fig. 3a–c and Supporting Information Table S1 and Fig. S9, $P < 0.05$ FWE-corrected at the cluster level).

Task-Based Functional Connectivity Results

The PPI analysis revealed reliable context-dependent functional connectivity for dACC as the seed region. In particular, the functional couplings of dACC (Fig. 4a–c) with posterior cingulate cortex (PCC, $x/y/z = -16/-20/56$ mm, $P < 0.05$ FWE-corrected at the cluster level) and amygdala-oribital frontal cortex (amygdala-OFC, $x/y/z = 20/24/-14$ mm, $P < 0.05$ FWE-corrected at the cluster level) were modulated by the Context × Split interaction (Alone [Unfair – Fair] > Group [Unfair – Fair]).

Among these functional couplings, the strength of the dACC-PCC connectivity showed a positive correlation with behavioral changes in amounts of punishment (i.e., diffusion of responsibility, $r = 0.52$, $P < 0.05$, Bonferroni corrected, Fig. 4d).

RSFC Results

Univariate correlational analysis

To explore whether behavioral responsibility diffusion was associated with intrinsic organization of brain networks, we next performed a voxel-wise correlation analysis between the average time series of the dACC and right putamen as seed regions and those of all of the other voxels in the brain. The functional connectivity strength of each voxel with the dACC/right putamen was further correlated with changes in punishment and responsibility across participants. The correlation analysis revealed that the strength of functional connectivity between right putamen and the following regions was negatively correlated with participants' changes in punishment: right dlPFC ($x/y/z = 22/44/50$ mm) and left TPJ ($x/y/z = -36/-74/34$ mm; Fig. 5a–c, $P < 0.05$ FWE-corrected at the cluster level). That is, stronger intrinsic functional connectivity between right putamen and these regions was associated with decreased responsibility diffusion in punishment.

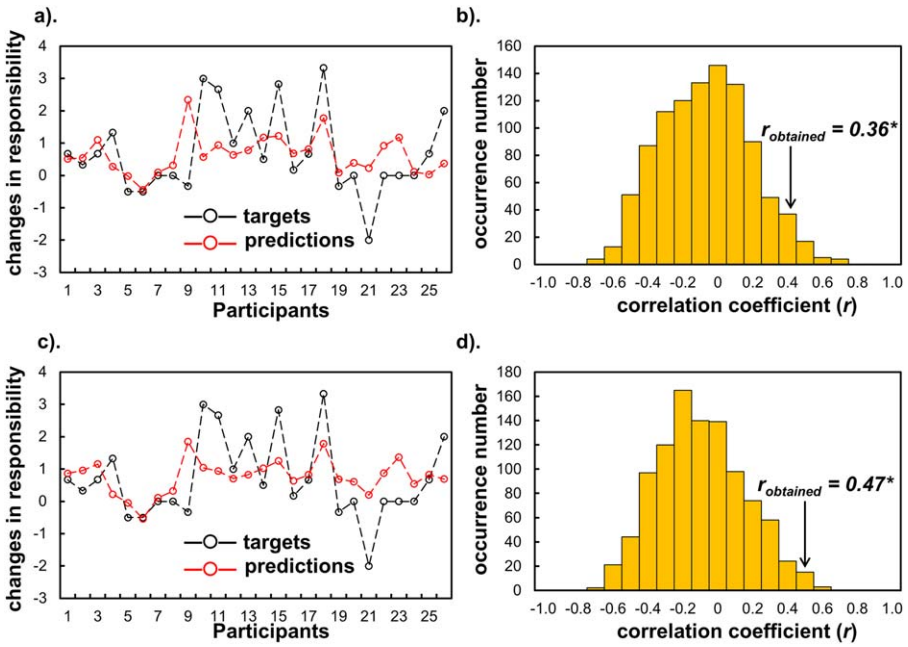

Figure 6.

MKL findings of the pattern regression analysis in predicting changes of responsibility ratings. **(a)** Line plot showing consistency between actual and predicted changes of responsibility ratings for the leave-one-subject-out cross validation procedures. **(b)** Permutation distribution of the correlation coefficient for the leave-one-subject-out cross validation procedures. **(c)** Line plot showing consistency between actual and predicted changes of responsibility ratings for the 10-fold cross validation procedures. **(d)** Permutation distribution of the correlation coefficient for the 10-fold cross validation procedures.*$P < 0.05$, FDR corrected. [Color figure can be viewed at wileyonlinelibrary.com]

Further, the strength of functional connectivity between right putamen and the mPFC/dmPFC ($x/y/z = 0/46/16$; $-2/54/50$ mm; Fig. 5d–f, $P < 0.05$ FWE-corrected at the cluster level) was negatively correlated with participants' changes in responsibility. That is, stronger intrinsic functional connectivity between right putamen and mPFC/dmPFC was associated with decreased changes in responsibility ratings. No significant correlations with responsibility diffusion were identified for the intrinsic functional connectivity with the dACC as a seed region.

Pattern regression analysis

The analysis aimed to complement the univariate correlational analysis by examining whether patterns of intrinsic connectivity with putamen or dACC could be used to decode individual differences in changes of punishment and responsibility ratings. Based on patterns of putamen- and dACC-related connectivity, the correlation coefficient (r) between actual and predicted changes in responsibility ratings was significant in both LOSOCV procedures ($r = 0.36$, $P < 0.05$, FDR corrected; Fig. 6a,b) and 10-fold cross-validation procedures ($r = 0.47$, $P < 0.05$, FDR corrected; Fig. 6c,d). Typical regions that contributed to the final decision model in the MKL included mPFC, insula, inferior frontal gyrus, and temporal pole among other regions (Supporting Information Table S6 and Fig. S11). For changes in amounts of punishment, no significant results were identified.

DISCUSSION

Given the critical role of altruistic punishment in maintaining widespread human cooperation, the mechanism underlying the phenomenon has long been a fundamental topic in psychology, economics, evolutionary biology, and neuroscience [Boyd et al., 2003; De Quervain et al., 2004; Fehr and Fischbacher, 2004a; Fehr and Gächter, 2002; Fowler, 2005]. Combining event-related fMRI with both task-based and RSFC, the current study explored the neural signatures underlying heterogeneity with respect to altruistic punishment across contexts and people. We demonstrated that impartial third-party decision-makers reduced their punishment of norm violations in the presence of other punishers, an effect mediated by diffusion of responsibility. Underlying these behavioral effects were attenuated neural responses of dACC and putamen, regions implicated in detecting norm violations. Further, we identified the context-dependent functional connectivity of dACC with PCC and amygdala/OFC, regions associated with reward processing. Regarding person-dependent

altruistic punishment, we found that the intrinsic functional connectivity of the putamen with brain regions critical to mentalizing and information integration—including TPJ, mPFC/dmPFC and dlPFC—contributed to individual differences in responsibility diffusion. Further, multivariate patterns of intrinsic putament- and dACC-related connectivity were sufficient to decode the responsibility ratings at the individual level. Together, our work provides important information for understanding the neural basis of contextual and individual variation in human altruistic punishment.

We first demonstrated that participants in the role of an unaffected third-party punished norm violations at the expense of perceived personal costs, thus providing evidence for human altruistic punishment [Fehr and Gächter, 2002]. This act of strong reciprocity is a hallmark of human civilization that contributes to the reinforcement of social norms (e.g., fairness) to sustain cooperation among genetically unrelated individuals [Fehr and Fischbacher, 2003, 2004b; Fehr and Gächter, 2002; Henrich et al., 2006]. The altruistic punishment recruited several brain regions, including dACC, AI, midbrain, dlPFC and inferior parietal lobule, which might reflect dynamic cognitive-affective-motivational processes that drive this prosocial behavior [Baumgartner et al., 2012; Feng et al., 2015b; Sanfey and Chang, 2008; Strobel et al., 2011]. In line with our findings, numerous studies have shown the involvement of similar neural circuits in altruistic punishment, which has been identified across human societies [De Quervain et al., 2004; Fehr and Gächter, 2002; Feng et al., 2015b,; Henrich et al., 2006; Strobel et al., 2011].

Further, our data revealed that human altruistic punishment is attenuated by the presence of others due to diffusion of responsibility. That is to say, people who felt less responsible in response to norm violations would punish less severely in the presence of others. These findings echo the assertion that responsibility diffusion plays a key role in attenuating altruistic behaviors in the presence of others [Hutcheson et al., 2015; Mynatt and Sherman, 1975; Rosenblatt et al., 1989]. At the neural level, the dACC and putamen, regions important in the detection and resolution of norm violations, exhibited attenuated responses in the presence of others compared with being alone. The dACC and putamen have been repeatedly implicated in encoding the deviations from normative expectations during social interactions, and their activations are predictive of altruistic punishment to norm violations [Chang and Sanfey, 2013; Civai, 2013; Gabay et al., 2014; Xiang et al., 2013]. For instance, recent studies employing computational modeling have highlighted the role of dACC in tracking social expectation violations that are closely associated with altruistic punishment [Chang and Sanfey, 2013; Xiang et al., 2013]. Likewise, the activity of putamen predicted punishment decisions by encoding the salience of norm violations [Gabay et al., 2014; Hu et al., 2015]. Therefore, our neuroimaging findings suggest that the presence of others renders norm violation a less salient event during social interactions.

Our task-based functional connectivity findings next uncovered how dACC worked together with other brain regions to determine the influence of responsibility diffusion. Compared with the presence of others, there was stronger connectivity of dACC with PCC and amygdala/OFC in response to norm violations while making decisions alone. Among other functions, the PCC has been implicated in reward-related processing [Ballard and Knutson, 2009; Levy and Glimcher, 2011; McClure et al., 2007]. The PCC responds both to receipt and to anticipation of reward and its activity is correlated with reward size [McCoy et al., 2003]. Moreover, the PCC is associated with social rewards elicited by cooperating with others [Suzuki et al., 2011] or receiving fair offers from others [Feng et al., 2015b]. Similarly, the amygdala/OFC has been implicated in encoding predictive reward values during decision-making [Gottfried et al., 2003; Schoenbaum et al., 1998]. Therefore, the current connectivity patterns suggest that in the alone context participants anticipated stronger satisfaction as a result of the punishment of norm violations. This conjecture is consistent with prior evidence showing that altruistic punishment recruits the engagement of reward-related processing due to the learned contingencies between norm-enforcing behavior and social rewards [De Quervain et al., 2004; Strobel et al., 2011]. Our data provide further support to this account by showing that stronger strength of dACC-PCC connectivity was associated with lower responsibility diffusion in altruistic punishment.

Finally, our RSFC findings demonstrated that the putamen functions in concert with brain regions associated with mentalizing (e.g., TPJ and mPFC/dmPFC) and information-integration (e.g., dlPFC) during the resting state to underlie the individual differences in responsibility diffusion. Among these regions, the TPJ and mPFC/dmPFC have been long regarded as important regions implicated in social cognition, particularly with respect to inferring others' intentions and motivations (i.e., mentalizing) [Frith and Frith, 2006; Lieberman, 2007; Rilling and Sanfey, 2011]. In complex social contexts, flexible and adaptive decision-making frequently requires predicting the actions of other people based on their intentions or desires. Accordingly, people's decisions about altruistic punishment often recruit TPJ and mPFC/dmPFC, presumably reflecting the need to model the minds of others [Baumgartner et al., 2012, 2014; Feng et al., 2016; Strobel et al., 2011]. Further, the dlPFC is associated with integrating context-dependent information (e.g., the presence of others) and converting them into actual punishment decisions [Feng et al., 2015b; Krueger and Hoffman, 2016]. For instance, disruption of the dlPFC diminishes altruistic punishment to norm violators [Knoch et al., 2006, 2008], presumably by impairing the integration of more abstract value (i.e., social norms) into decision-making [Buckholtz

and Marois, 2012]. Together, these findings revealed that brain regions frequently engaged by altruistic punishment exhibited intrinsic functional organization in the absence of an explicit task, further elucidating possible mechanisms of individual differences in effects of responsibility diffusion on altruistic punishment. These findings echo the intrinsic perspective of brain function by complementing previous observations of the predictive significance of spontaneous activity with respect to cognition and behavior [Fox et al., 2007; Wei et al., 2012] and personality traits [Cox et al., 2012; Kong et al., 2015; Xiang et al., 2016].

In line with these findings, previous studies have demonstrated that altruistic behaviors are modulated by a variety of personal attributes, including self-reported personality and degrees of empathic responding measured at the neural level [Böckler et al., 2017; Hein et al., 2010; Mathur et al., 2010; Tusche et al., 2016]. For instance, empathic neural responses to others predict altruistic behaviors orientated to both the empathized targets in specific [Hein et al., 2010; Masten et al., 2011; Mathur et al., 2010] and daily life [Morelli et al., 2014; Rameson et al., 2012]. Whereas previous findings have provided insights on the contributions of personality and empathy-related neural responses to individual variations in altruistic behaviors, our results involving spontaneous neuronal fluctuations open a new avenue for understanding the neural mechanism of individual differences in altruistic behaviors.

Several potential limitations related to the current study should be acknowledged. First, it is possible that the dACC and putamen revealed in the activation analysis reflect differences in task difficulty between conditions. According to this interpretation, however, one would expect stronger activations of these regions in the group than alone context, because the group context included more players and elements to consider for a decision. Instead, this study revealed that the activity of dACC and putamen were attenuated in the group context, which was contradictory to the account of task difficulty. Second, the current findings of RSFC-behavior relationship should be considered with caution due to the moderate reliability of RSFC [Cao et al., 2014; Telesford et al., 2010]. Although an increasing body of research has employed the RSFC approach to explore the neural underpinnings of individual differences [Dubois and Adolphs, 2016; Finn et al., 2015], the reliability of the associations between RSFC and cognitive functions/behaviors awaits further investigation.

In summary, our findings revealed the neural basis underlying the context- and person-dependent altruistic punishment. The presence of putative others resulted in a diffusion of responsibility that mediated a reduction in punishment to social norm violations. These behavioral effects paralleled neural responses in the dACC and putamen, as well as their functional connectivity with areas implicated in reward-related processing (PCC, amygdala/OFC). Finally, we demonstrated that intrinsic functional connectivity between putamen and areas associated with mentalizing (e.g., TPJ and mPFC/dmPFC) and information-integration (e.g., dlPFC) contributed to individual variation in responsibility diffusion. Our findings have significant implications for disciplines studying human altruism and provide potential neurocognitive mechanisms for heterogeneous altruistic behaviors across both social contexts and individuals.

ACKNOWLEDGMENTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

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

Batson CD, Bolen MH, Cross JA, Neuringer-Benefiel HE (1986): Where is the altruism in the altruistic personality? J Pers Soc Psychol 50:212.

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, Schiller B, Rieskamp J, Gianotti LR, Knoch D (2014): Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. Soc Cogn Affect Neurosci 9:653–660.

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

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

Birn RM (2007): The behavioral significance of spontaneous fluctuations in brain activity. Neuron 56:8–9.

Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS (1995): Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magn Reson Med 34:537–541.

Boyd R, Gintis H, Bowles S, Richerson PJ (2003): The evolution of altruistic punishment. Proc Natl Acad Sci 100: 3531–3535.

Böckler A, Sharifi M, Kanske P, Dziobek I, Singer T (2017): Social decision making in narcissism: Reduced generosity and increased retaliation are driven by alterations in perspective-taking and anger. Pers Individ Differences 104:1–7.

Buckholtz JW, Asplund CL, Dux PE, Zald DH, Gore JC, Jones OD, Marois R (2008): The neural correlates of third-party punishment. Neuron 60:930–940.

Buckholtz JW, Marois R (2012): The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. Nat Neurosci 15:655–661.

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

Cao H, Plichta MM, Schäfer A, Haddad L, Grimm O, Schneider M, Esslinger C, Kirsch P, Meyer-Lindenberg A, Tost H (2014): Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. Neuroimage 84:888–900.

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

Chekroun P, Brauer M (2002): The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations. Eur J Soc Psychol 32: 853–867.

Civai C (2013): Rejecting unfairness: emotion-driven reaction or cognitive heuristic?. Frontiers in human Neuroscience 7:126.

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. Soc Cogn Affect Neurosci 10:1054–1060.

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 Cogn Affect Neurosci 8:424–431.

Cox CL, Uddin LQ, Di Martino A, Castellanos FX, Milham MP, Kelly C (2012): The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. Soc Cogn Affect Neurosci 7: 727–737.

Crişan LG, Pană S, Vulturar R, Heilman RM, Szekely R, Drugă B, Dragoş N, Miu AC (2009): Genetic contributions of the serotonin transporter to social learning of fear and economic decision making. Soc Cogn Affect Neurosci 4:399–408.

Cui Z, Xia Z, Su M, Shu H, Gong G (2016): Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. Hum Brain Mapp 37:1443–1458.

De Quervain DJ, Fischbacher U, Treyer V, Schellhammer M (2004): The neural basis of altruistic punishment. Science 305:1254.

Dubois J, Adolphs R (2016): Building a science of individual differences from fMRI. Trends Cogn Sci 20:425–443.

Eisenberg N, Miller PA (1987): The relation of empathy to prosocial and related behaviors. Psychol Bull 101:91.

Eklund A, Nichols TE, Knutsson H (2016): Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci USA 113:7900–7905.

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

Fehr E, Bernhard H, Rockenbach B (2008): Egalitarianism in young children. Nature 454:1079–1083.

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

Fehr E, Fischbacher U (2004a): Social norms and human cooperation. Trends Cogn 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 (1999): Cooperation and punishment in public goods experiments. Am Econ Rev 90:980–994.

Fehr E, Gächter S (2000): Cooperation and Punishment in Public Goods Experiments. Am Econ Rev 90:980–994.

Fehr E, Gächter S (2002): Altruistic punishment in humans. Nature 415:137–140.

Feng C, Azarian B, Ma Y, Feng X, Wang L, Luo YJ, Krueger F (2017): Mortality salience reduces the discrimination between in-group and out-group interactions: A functional MRI investigation using multi-voxel pattern analysis. Hum Brain Mapp 38:1281–1298.

Feng C, Deshpande G, Liu C, Gu R, Luo YJ, Krueger F (2016): Diffusion of responsibility attenuates altruistic punishment: A functional magnetic resonance imaging effective connectivity study. Hum Brain Mapp 37:663–677.

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

Feng C, Lori A, Waldman ID, Binder EB, Haroon E, Rilling J (2015a): A common oxytocin receptor gene (OXTR) polymorphism modulates intranasal oxytocin effects on the neural response to social cooperation in humans. Genes Brain Behav 14:516–525.

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

Fernandes O, Portugal LC, Rita de Cássia SA, Arruda-Sanchez T, Rao A, Volchan E, Pereira M, Oliveira L, Mourao-Miranda J (2017): Decoding negative affect personality trait from patterns of brain activation to threat stimuli. Neuroimage 145:337–345.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT (2015): Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. Nat Neurosci 18:1664–1671.

Fischbacher U, Gächter S, Fehr E (2001): Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett 71:397–404.

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.

Fowler JH (2005): Altruistic punishment and the origin of cooperation. Proc Natl Acad Sci USA 102:7047–7049.

Fox MD, Snyder AZ, Vincent JL, Raichle ME (2007): Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. Neuron 56:171–184.

Friston KJ, Buechel C, Fink G, Morris J, Rolls E, Dolan R (1997): Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Frith CD, Frith U (2006): The neural basis of mentalizing. Neuron 50:531–534.

Gabay AS, Radua J, Kempton MJ, Mehta MA (2014): The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 47:549–558.

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003): Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19: 200–207.

Gottfried JA, O'Doherty J, Dolan RJ (2003): Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104–1107.

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.

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. Soc Cogn Affect Neurosci 9: 1498–1505.

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

Harmelech T, Malach R (2013): Neurocognitive biases and the patterns of spontaneous correlations in the human cortex. Trends Cogn Sci 17:606–615.

Haruno M, Frith CD (2010): Activity in the amygdala elicited by unfair divisions predicts social value orientation. Nat Neurosci 13:160–161.

Haruno M, Kimura M, Frith CD (2014): Activity in the Nucleus Accumbens and Amygdala Underlies Individual Differences in Prosocial and Individualistic Economic Choices. J Cogn Neurosci 26:1861–1870.

Hein G, Silani G, Preuschoff K, Batson CD, Singer T (2010): Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. Neuron 68: 149–160.

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. Social Cognitive and Affective Neuroscience nsv086.

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. Human 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 Bus 59:S285–S300.

Knoch D, Nitsche MA, Fischbacher U, Eisenegger C, Pascual-Leone A, Fehr E (2008): Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. Cereb Cortex 18:1987–1990.

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

Kong F, Hu S, Wang X, Song Y, Liu J (2015): Neural correlates of the happy life: The amplitude of spontaneous low frequency fluctuations predicts subjective well-being. Neuroimage 107:136–145.

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, Hoffman M (2016): The emerging neuroscience of third-party punishment. Trends Neurosci 39:499–501.

Latané B, Darley JM (1968): Group inhibition of bystander intervention in emergencies. J Pers Soc Psychol 10:215–221.

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.

Lewis CM, Baldassarre A, Committeri G, Romani GL, Corbetta M (2009): Learning sculpts the spontaneous activity of the resting human brain. Proc Natl Acad Sci USA 106:17558–17563.

Lieberman MD (2007): Social cognitive neuroscience: A review of core processes. Annu Rev Psychol 58:259–289.

Martin-Raugh MP, Kell HJ, Motowidlo SJ (2016): Prosocial knowledge mediates effects of agreeableness and emotional intelligence on prosocial behavior. Pers Individ Differences 90:41–49.

Masten CL, Morelli SA, Eisenberger NI (2011): An fMRI investigation of empathy for 'social pain'and subsequent prosocial behavior. Neuroimage 55:381–388.

Mathur VA, Harada T, Lipke T, Chiao JY (2010): Neural basis of extraordinary empathy and altruistic motivation. Neuroimage 51:1468–1475.

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

McCoy AN, Crowley JC, Haghighian G, Dean HL, Platt ML (2003): Saccade reward signals in posterior cingulate cortex. Neuron 40:1031–1040.

McDermott R, Tingley D, Cowden J, Frazzetto G, Johnson DD (2009): Monoamine oxidase A gene (MAOA) predicts behavioral aggression following provocation. Proc Natl Acad Sci USA 106:2118–2123.

McLaren DG, Ries ML, Xu G, Johnson SC (2012): A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. Neuroimage 61: 1277–1286.

Mennes M, Kelly C, Zuo X-N, Di Martino A, Biswal BB, Castellanos FX, Milham MP (2010): Inter-individual differences in resting-state functional connectivity predict task-induced BOLD activity. Neuroimage 50:1690–1701.

Morelli SA, Rameson LT, Lieberman MD (2014): The neural components of empathy: Predicting daily prosocial behavior. Soc Cogn Affect Neurosci 9:39–47.

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

Oda R, Machii W, Takagi S, Kato Y, Takeda M, Kiyonari T, Fukukawa Y, Hiraishi K (2014): Personality and altruism in daily life. Pers Individ Differences 56:206–209.

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

Poldrack RA, Mumford JA, Nichols TE (2011): Handbook of functional MRI data analysis. New York: Cambridge University Press.

Raichle ME (2006): The brain's dark energy. Science 314:1249.

Raichle ME (2011): The restless brain. Brain Connect 1:3–12.

Raichle ME (2015): The brain's default mode network. Annu Rev Neurosci 38:433–447.

Rameson LT, Morelli SA, Lieberman MD (2012): The neural correlates of empathy: Experience, automaticity, and prosocial behavior. J Cogn Neurosci 24:235–245.

Rilling JK, Sanfey AG (2011): The neuroscience of social decision-making. Annu Rev Psychol 62:23–48.

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 Pers Soc Psychol 57:681.

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

Schoenbaum G, Chiba AA, Gallagher M (1998): Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci 1.

Schrouff J, Monteiro J, Joao Rosa M, Portugal L, Phillips C, Mourao-Miranda J. (2014): Can we interpret linear kernel machine learning models using anatomically labelled regions? Poster presented at the 20th Annual Meeting of the Organization for Human Brain Mapping Hamburg, Germany.

Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Richiardi J, Mourão-Miranda J (2013): PRoNTo: Pattern recognition for neuroimaging toolbox. Neuroinformatics 11:319–337.

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.

Suzuki S, Niki K, Fujisaki S, Akiyama E (2011): Neural basis of conditional cooperation. Soc Cogn Affect Neurosci 6:338–347.

Taubert M, Lohmann G, Margulies DS, Villringer A, Ragert P (2011): Long-term effects of motor training on resting-state networks and underlying brain structure. Neuroimage 57:1492–1498.

Telesford QK, Morgan AR, Hayasaka S, Simpson SL, Barret W, Kraft RA, Mozolic JL, Laurienti PJ (2010): Reproducibility of graph metrics in fMRI networks. Front Neuroinformatics 4.

Tusche A, Böckler A, Kanske P, Trautwein F-M, Singer T (2016): Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. J Neurosci 36:4719–4732.

van Bommel M, van Prooijen J-W, Elffers H, Van Lange PA (2012): Be aware to care: Public self-awareness leads to a reversal of the bystander effect. J Exp Soc Psychol 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 71–91. Cambridge, MA: MIT Press.

Wang X, Han Z, He Y, Caramazza A, Song L, Bi Y (2013): Where color rests: spontaneous brain activity of bilateral fusiform and lingual regions predicts object color knowledge performance. Neuroimage 76:252–263.

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

Wei T, Liang X, He Y, Zang Y, Han Z, Caramazza A, Bi Y (2012): Predicting conceptual processing capacity from spontaneous neuronal activity of the left middle temporal gyrus. J Neurosci 32:481–489.

Wiesenthal DL, Austrom D, Silverman I (1983): Diffusion of Responsibility in Charitable Donations. Basic Appl Soc Psychol 4:17–27.

Woo C-W, Krishnan A, Wager TD (2014): Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. Neuroimage 91:412–419.

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

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

Xiang Y, Kong F, Wen X, Wu Q, Mo L (2016): Neural correlates of envy: Regional homogeneity of resting-state brain activity predicts dispositional envy. Neuroimage 142:225–230.