Behavioural Brain Research 332 (2017) 136–144

Contents lists available at ScienceDirect

Behavioural Brain Research

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

The neural correlates of reciprocity are sensitive to prior experience of reciprocity

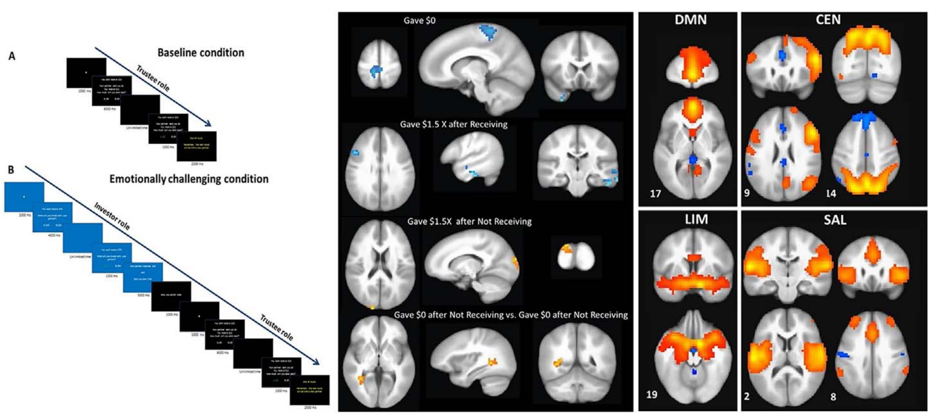
Ricardo Cáceda[a,*], Stefania Prendes-Alvarez[b], Jung-Jiin Hsu[c], Shanti P. Tripathi[a], Clint D. Kilts[a], G. Andrew James[a]

[a] *Brain Imaging Research Center, Psychiatric Research Institute, University of Arkansas for Medical Sciences,4301 West Markham St., Slot #554, Little Rock, AR 72205, United States*

[b] *Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th Street, Miami, FL 33136, United States*

[c] *Department of Radiology and Biomedical Engineering, University of California San Francisco School of Medicine, 505 Parnassus Avenue, M-391, San Francisco, CA 94143-0628, United States*

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Reciprocity
Prosocial behavior
Selfishness
Trust game
Independent component analysis
Neural networks

ABSTRACT

Reciprocity is central to human relationships and is strongly influenced by multiple factors including the nature of social exchanges and their attendant emotional reactions. Despite recent advances in the field, the neural processes involved in this modulation of reciprocal behavior by ongoing social interaction are poorly understood. We hypothesized that activity within a discrete set of neural networks including a putative moral cognitive neural network is associated with reciprocity behavior. Nineteen healthy adults underwent functional magnetic resonance imaging scanning while playing the trustee role in the Trust Game. Personality traits and moral development were assessed. Independent component analysis was used to identify task-related functional brain networks and assess their relationship to behavior. The saliency network (insula and anterior cingulate) was positively correlated with reciprocity behavior. A consistent array of brain regions supports the engagement of emotional, self-referential and planning processes during social reciprocity behavior.

* Corresponding author.
E-mail addresses: RCaceda@uams.edu (R. Cáceda), sprendes@med.miami.edu (S. Prendes-Alvarez), jung-jiin.hsu@ucsf.edu (J.-J. Hsu), CDKilts@uams.edu (C.D. Kilts), GAJames@uams.edu (G.A. James).

http://dx.doi.org/10.1016/j.bbr.2017.05.030
Received 3 March 2017; Received in revised form 3 May 2017; Accepted 10 May 2017
Available online 24 May 2017
0166-4328/ Published by Elsevier B.V.

1. Introduction

Social reciprocity, defined as behavioral response to a positive action with another positive action, is central to an individual's development of healthy interpersonal relationships, learning of social norms and normative ethics, and integration into society [1]. Reciprocity and other prosocial behaviors, such as altruism, fairness, and cooperation, are adaptive behaviors common to many cultures and religious traditions. Human subjects engage in prosocial behaviors even at their own expense [2], and thus reciprocity can overlap with altruistic behavior. However, reciprocity and other prosocial behaviors are multi-determined and dissociable products of complex and dynamic transactional processes [3]. For instance, reciprocity can be seen as self-serving, whether by promoting the benefit of future reciprocity, increasing one's social status or reputation, resolving the conflict of seeing somebody else suffer, enabling redemption or release from guilt, or even by increasing reproductive population fitness [2,4–7].

The pervasive human distinction between "us" and "others" is traditionally based on kinships but tends to be determined by social proximity and contact. Personal interaction with others makes decision making less utilitarian and more emotionally driven [8]. For example, when research subjects believe they are interacting with other humans (rather than computers) neuroimaging has shown greater engagement of the ventral striatum and ventromedial prefrontal cortex (VMPFC), brain regions that are both associated with reward processing [9,10]. However, the neural processes underlying the modulating effects of social interaction, either positive or negative, on basic reciprocal behavior are largely unknown.

Monetary exchange paradigms have yielded valuable insight into the neurobiology of cooperation [11], trust [12,13], and agency [14]. In the present study, we sought to further characterize the neural mechanisms underlying individual variation in reciprocity behavior during a monetary decision making task, the Trust Game. Variations in task conditions allowed to assess the behavioral and neural processing correlates of reciprocating a gift before and after a social exchange in which a gift to another was or was not reciprocated. We hypothesized that reciprocity correlates with the engagement of neural processing networks including a putative moral cognitive neural network composed of the frontal pole, anterior cingulate cortex (ACC), superior temporal cortex (STS), precuneus, and posterior cingulate cortex (PCC) [15–17].

2. Materials and methods

2.1. Participants

We recruited 30 participants aged 18–30 years, 12 men and 18 women. The racially and ethnically diverse sample consisted of 15 Caucasians, 13 Hispanics, and 2 African Americans. Twenty-four participants were college graduate or undergraduate students, five were employed, and one participant was unemployed. Eleven participants were excluded from the brain imaging analysis due to head motion artifact (n = 1) or for not varying their behavioral response (always reciprocating), which precluded modeling different response types (n = 10). There were no participants who always failed to reciprocate. Demographics of the 19 subjects included in the imaging data analysis are provided in Table 1.

Inclusion criterion was: a) age 18–30. Exclusion criteria were: a) neurological or psychiatric history, including active substance use, other than tobacco; b) inability to write and speak English; c) inability to provide written informed consent; d) non-removable ferromagnetic objects; and e) history of claustrophobia.

Subjects provided written informed consent to participate in accordance with the Institutional Review Board of the University of Miami. Next, participants were assessed for personality trait composition using the NEO Five-Factor Inventory (NEO-FFI), moral reasoning development with the Defining Issues Test-2 (DIT-2), trait impulsivity with the Barratt Impulsiveness Scale-11 (BIS-11), and history of childhood maltreatment with the Childhood Trauma Questionnaire (CTQ). Lastly, participants underwent functional magnetic resonance imagining (fMRI) scanning while playing the trustee role in the Trust Game (described below).

Table 1
Descriptive demographics and functional level of subjects included in imaging analysis.

Demographic variables	
Number of subjects	19
Age	24.7 ± 0.5 [a]
Gender (F/ M)	12/7
Race (AA/CA/H)	2/13/4
Education (years)	16.8 ± 0.8 [a]
Handedness (R/ L)	19/0
Occupation (student/work/unemployed)	15/3/1

[a] Mean ± standard deviation.

2.2. Psychological assessment scales

The NEO-FFI is an abbreviated version of the NEO Personality Inventory, which assesses five personality domains: neuroticism, extraversion, openness, agreeableness and conscientiousness [18].

The Defining Issues Test (DIT-2) activates and assesses moral schemas in terms of "importance judgments". Based on Kohlberg's theory of moral development, the subject's task is to read a moral dilemma, then rate and rank corresponding statements in terms of their moral importance. Measures for the following schemas are obtained: a) Personal Interest, the least developmentally advanced, focuses on a self-centered utilitarian approach; b) Maintaining Norms, emphasizes behavior driven by rules; and c) Postconventional, the most developmentally advanced moral stage, in which laws are not simply blindly accepted, but are scrutinized in order to ensure society-wide benefit [19]. Completed DIT-2 assessments were sent to the Center for Ethical Development [19] for automated scoring.

The BIS-11 measures three aspects of impulsivity: a) attentional/cognitive impulsivity denotes a lack of cognitive persistence with an inability to tolerate cognitive complexity; b) motor impulsivity signifies a tendency to act impulsively; and c) non-planning impulsivity indicates a lack of sense of the future [20].

2.3. MRI acquisition

Participants underwent brain imaging using a Siemens 3T Trio MRI scanner at the University of Miami, Applebaum MRI Center. The MRI session included the following scans: a magnetization prepared rapid acquisition gradient echo (MPRAGE) T1 anatomic scan (5 min), a resting-state (rs) echo-planar imaging (EPI) fMRI scan (rs-fMRI, 5 min) and an echo-planar imaging (EPI) fMRI scan during the Trust Game (15 min, see below). The MPRAGE scan had the following parameters: matrix = 248 × 256, 220 sagittal slices, TR/TE/FA = 2300 ms/3.08 ms/9°, final resolution = 0.86 × 0.86 × 1.20 mm^3 resolution. The Trust Game was conducted across three fMRI scans with the following parameters: TR/TE/FA = 2000 ms/30 ms/90°, FOV = 220 × 220 mm, matrix = 64 × 64, 25 axial slices (acquired parallel to the AC-PC line with interleaved slice acquisition), slice thickness = 5 mm, final resolution 3.44 × 3.44 × 5.00 mm^3, 154 images for 5 min 8 s. DTI and rs-fMRI scans were not analyzed for the current work and are described elsewhere [21].

The Trust Game was conducted across three fMRI runs within a single session. Prior to scanning, participants received instructions for playing a variation of the Trust Game (described below) and played a practice game to ensure they understood the instructions. Next, participants underwent fMRI while playing the Trust Game. During scanning, participants viewed task stimuli and instructions via back-projection and responded via MRI-compatible keypad. After scanning, subjects were debriefed to determine: a) their decision making strategy, b) whether they believed they were playing against other people, and c) their emotional response when the other person did not reciprocate in the emotionally challenging condition. Participants were then debriefed about the game's deception (i.e., that they had been playing against a computer and not other players).

2.3. The Trust Game is a two-person reciprocal exchange game widely used to model interpersonal trust and reciprocity [11], which has traditionally been used to assess the trust of the first player or trustee. We used a modified version in which the participant decides whether or not to reciprocate the other player's trust [22]. We used two conditions in the game, each consisting of 30 trials during which the participant played the Trust Game once with 30 different anonymous individuals. In actuality these answers were pre-generated to span offers within the $6- $80 range. Participants were instructed that they would play the Trust Game with other people who have already played the game and that their responses (offers) had been recorded and paired with those of the participants themselves. They were also informed that their and the other players' monetary compensation would be determined by the outcome from a randomly chosen trial. This was to convey to the participants that the investment sums had been drawn from a distribution of amounts actually proposed by real players, so participants believe they were exposed to realistic offers. The first condition (Baseline) was presented during Run 1. The second condition, Emotionally Challenging (EC), was presented during two runs (Runs 2 and 3) with 15 rounds per run. For Baseline trials, the participant (the trustee) and the other anonymous individual (the investor) were endowed with an initial amount of money that ranged from $6- $80, mean $37.8, of which the investor then gave a certain amount (25%-75%) of the money ($X) to the participant. Next, the participant (trustee) received three times what the investor gave ($3X), and was asked to choose how much money he/she wanted to give back to the investor, according to a forced choice of either $1.5X or $0. The timing of the stimuli in each condition is shown in Fig. 1.

In each round of the EC conditions, participants first played the Investor role and immediately afterwards they played the Trustee role with the same player. The first part of the EC trials had the same structure as the Baseline condition, except that the participant played the role of the Investor, with two fixed options ($Y) to invest ranging between $3 and $50 or half this value (e.g. $6 or $3, or $ $40 or $20). By design, in 50% of the rounds participants received $1.5Y back from the other player (trustee) and in the other 50% participants received $0. The EC trials were designed to manipulate the tendency to reciprocate based on experiences of trusting versus non-trusting social interactions and their attendant pro-social and anti-social emotional reactions. Finally, the Baseline condition was repeated as a Test condition to assess the impact of these different trust experiences on reciprocity by the participant. The timing of the stimuli in each round is shown in Fig. 1. Upon completion of the experiments, patients were debriefed for the credibility of the manipulation, the rationale to make the decisions they made, and the deception regarding the computer-generated partners was revealed. Lastly, all participants were compensated $20 for their participation in the study. Additionally, participants received the dollar amount of one of their answers which was chosen randomly ($6-$80).

We measured reciprocity, altruism and selfish behavior as follows. **Reciprocity** was operationally defined by the number of trials in which participants gave money back in the Baseline condition. In the Test condition, **altruism** was operationally defined as the number of times participants gave money back despite not receiving any from the other player in the EC condition. **Selfish** behavior was operationally measured by the number of trials in which participants did not give money back after having received money from the other player in the EC condition.

We used binary logistic model to estimate variables affecting reciprocating behavior in the EC condition. The factor included in the model was reciprocating behavior of the Investor in the first round (yes or no). Reciprocity behavior in the Baseline condition was included as covariate.

2.4. fMRI preprocessing

Image processing and analysis were conducted with AFNI (National Institute of Mental Health) [23] and MATLAB (MathWorks). Anatomic data underwent skull stripping, spatial normalization to the icbm452 brain atlas, and segmentation into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) with FSL [24]. Functional data underwent despiking; slice correction; deobliquing (to $3 \times 3 \times 3$ mm^3 voxels); motion correction (using the 10th timepoint); transformation to the spatially normalized anatomic image; regression of nuisance signals including motion parameters, mean timecourse of subject's WM mask, and mean timecourse of subject's CSF; spatial smoothing with a 5-mm FWHM Gaussian kernel; and scaling to percent signal change.

2.5. fMRI data analysis

The fMRI data were analyzed using MATLAB and AFNI in a two-stage, random effects procedure. In the first stage, task-related changes in blood oxygen dependent (BOLD) activity were modeled for each subject using generalized linear modeling (GLM) via AFNI's 3dDeconvolve command with the standard canonical hemodynamic response function. The Baseline condition GLM modeled the regressors "Gave $X" and "Gave $0". The Test condition GLM modeled the regressors "Gave $1.5X After Receiving", "Gave $1.5X After Not Receiving", "Gave $0 After Receiving", and "Gave $0 After Not Receiving". Both GLMs included participant head motion as six nuisance parameters (lateral head motion: x, y, z; rotational head motion: roll, pitch, yaw). The second stage consisted of one-sample *t*-tests (if contrasting a condition versus rest) or two-sample *t*-tests (if contrasting two conditions) to test if task-related changes in BOLD were consistent for the entire sample. Planned contrasts for the Trust Game results included 1) Participant Gave $1.5X, 2) Participant Gave $0, and 3) general linear test contrast of Gave $1.5X vs Gave $0 under the Baseline condition; and 1) Participant Gave $1.5X After Not Receiving Money, 2) Gave $1.5X After Receiving Money, 3) Gave $0 After Receiving Money, 4) Gave $0 After Not Receiving Money, 5) the contrast Gave $0 After Receiving Money vs Gave $0 After Not Receiving Money, and 6) the contrast Gave $1.5X After Not Receiving Money vs. Gave $1.5X After Receiving in the Test or Emotionally Challenging condition. Four participants during the Baseline condition and seven participants during the Test condition always reciprocated; the contrasts of Gave $X/$1.5X versus Gave $0 could therefore not be modeled for these participants, so they were excluded from data analyses. Minimum cluster sizes were determined via a two-step process. Step 1 estimated the spatial smoothness of the data using the AFNI program 3dFWHMx, which was recently revised to model spatial smoothness as a non-Gaussian autocorrelation function (ACF) to adjust for inflated false-positive rates (Eklund, Nichols and Knutsson, 2016). 3dFWHMx calculated spatial smoothness as an ACF fitting the equation $ACF(r) = a * \exp(-r^2/2b^2) + (1-a)*\exp(-r/c)$, where r is the radius and a, b, and c are fitted parameters. The ACF was calculated using the residual error of the fit of that subject's fMRI data to the task-based GLM and subject-specific masks of the fMRI data. Median values for parameters a, b, c were calculated for each task (run1 and combined runs 2 and 3). Run 1 gave the more conservative values (median a = 0.6748, b = 6.9667, c = 14.9588 for estimated FWHMx = 16.45); these clustering parameters were used for both Run 1 and combined Runs 2 and 3. Using these parameters, AFNI's 3dClustSim estimated that clusters of significant task-related activity ($p < 0.05$ with family-wise correction) could be achieved by bi-sided thresholding (where positive and negative clusters are thresholded separately) at $p < 0.001$ (uncorrected) and minimum cluster size = 62

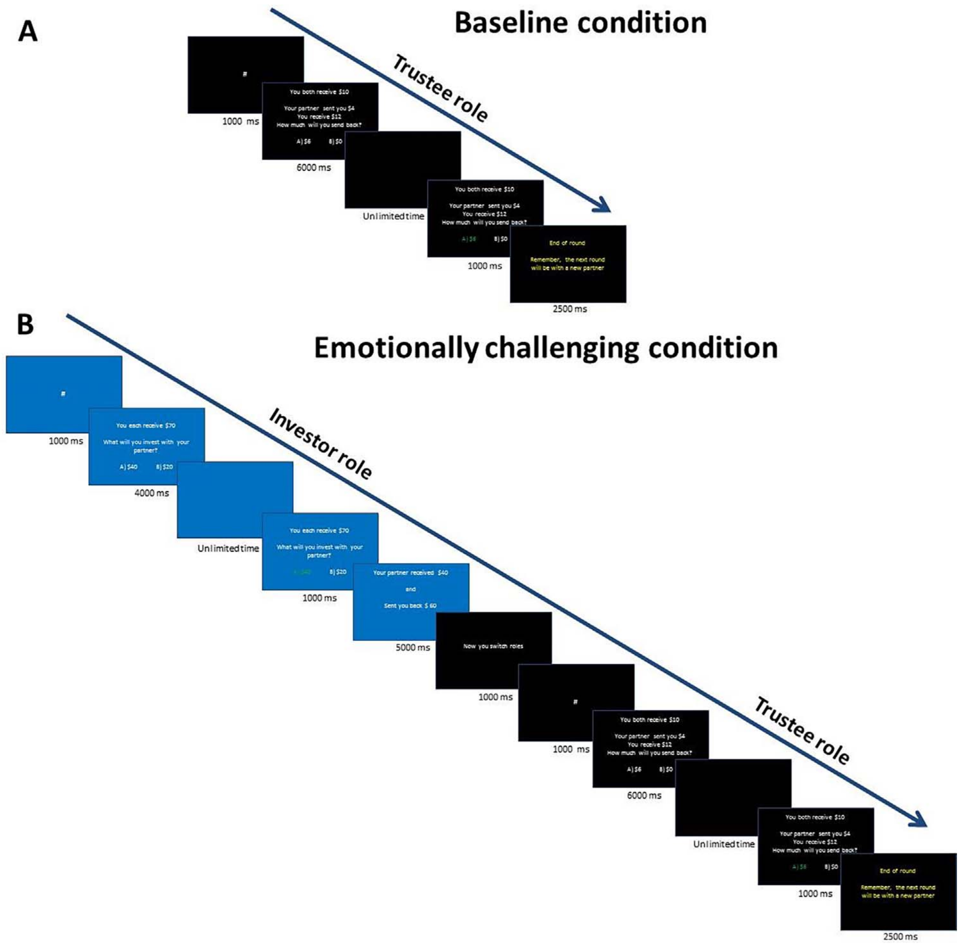


Fig. 1. Timeline for Baseline and Emotionally Challenging (EC) conditions in the Trust Game. A. Baseline condition, the participant plays the role of trustee (blue) who receives a variable amount of money equal to three times the sum X sent by the investor and needs to decide whether to give back some (1.5X) or nothing; B. Emotionally challenging (EC) condition, the participant (blue) plays the role of investor and chooses the amount of money he or she will give to the other player who either reciprocates ($1.5Y) or not ($0). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

voxels (1674 mm^3) with NN = 2 (voxels connected by contiguous faces or edges).

2.5.1. Multivariate fMRI data

After preprocessing, group-level independent component analysis (ICA) of functional imaging data sets was conducted using MATLAB and the Group ICA of fMRI Toolbox (GIFT v1.3; http://mialab.mrn.org/software/), an approach for blind-source separation of a complex mixture of signals into spatially and temporally distinct sources (independent components [IC]) [25]. ICA was run using Infomax algorithm to solve for 20 components (i.e. 20 brain networks) using Run 1 fMRI data (Baseline task). For consistent comparisons across tasks, the same components identified from the Baseline task were used to analyze the Emotionally Challenging task (Runs 2 and 3 data). The following options were used: back-reconstruction using GICA3, subject-specific principal component analysis using expectation maximization and stacked datasets, full storage of covariance matrix to double precision, usage of selective eigenvariate solvers, two-step data reduction with 40 principal components in the first step, and scaling to z-scores. ICA was repeated 20 times using the ICASSO algorithm to identify the most reliable and stable components across all iterations. The ICASSO stability indices (all $i_{Q} >$ 0.95) indicated a reliable solution using 20 components.

2.5.2. Regression analysis

Based on visual inspection and in comparison with the literature [26], we classified the independent components as follows. Nine components represented noise (such as ventricle fluctuations or head motion) and were excluded to reduce Type I error. Six components represented networks not typically associated with decision making and reward processing: one component represented sensorimotor, three representing the visual system, one representing the language system and one representing the cerebellum. These components were also omitted from subsequent analyses to reduce Type I error. All 20 components are depicted in the Supplementary Materials.

The six remaining components represented networks typically associated with hot and/or cold cognition belonging to the known brain networks: a) default-mode network (DMN) associated with self-related cognitive activity including autobiographical, self-monitoring and social cognitive functions, anchored in the mPFC, medial temporal lobe, precuneus, PCC, and TPJ [27] (IC 17); b) central executive network (CEN), linked to identification of relevant stimuli and execution of goal directed behaviors, anchored in the DLPFC and posterior parietal cortex (IC 9 and 14); c) the mesolimbic system network, associated with processing of emotional stimuli, anchored in the ventral tegmental area, striatum, amygdala and ventromedial PFC (LIM) (IC 19); and the saliency network (SAL), associated with monitoring salience if external inputs and internal events, anchored in the anterior cingulate and the anterior cingulate (IC 2 and 8). In order to focus on specialized networks underlying social decision making we reviewed *meta*-analytic data [28–31] that described active involvement of the salience network. For the two components belonging to the SAL, timeseries were calculated for each subject and fMRI run by back-projecting the ICA voxel-wise spatial β-map to each image timepoint, thus generating weighted timeseries of network activity for each subject and component. GLM identified task-related changes in these weighted timeseries for each subject, and two-tailed one-sample *t*-tests determined if group-level engagement of components during the contrasts significantly differed from 0. The Baseline task GLM had 3 contrasts (Participant Gave $1.5X, Participant Gave $0, and general linear test contrast of Gave $1.5X vs

Gave $0). The Emotionally Challenging task GLM used the −concat option to combine the two fMRI runs and had 6 GLM contrasts (Participant Gave $1.5X After Not Receiving Money, Gave $1.5X After Receiving Money, Gave $0 After Receiving Money, Gave $0 After Not Receiving Money, the contrast Gave $0 After Receiving Money vs Gave $0 After Not Receiving Money, and the contrast Gave $1.5X After Not Receiving Money vs. Gave $1.5X After Receiving Money). Each condition contrast underwent false discovery rate (FDR) correction for the number of components being studied (n = 2) at $q = 0.05$ using the Matlab FDR program [32].

Next, given the colinearity between some of our behavioral variables, we performed a multiple robust regression (using MATLAB's *robustfit* command with the Huber weighting function and tuning parameter 1.345) to relate behavioral measures to task-related recruitment of brain networks. We modeled behavioral meaures of personality dimensions, PI, PC, stage 4 and N2 subscores of DIT-2, and BIS-11 scores as independent variables, which were regressed to the dependent variable of task-related brain network activity (i.e. the GLM betas derived from deconvolution of ICA timeseries described above). Robust regression was chosen over standard linear regression for its greater resiliency to the effects of outliers, which are common in neuroimaging data [33].

3. Results

3.1. Behavioral outcomes

In the Baseline condition, the 19 participants included in the fMRI analysis reciprocated (Gave $1.5X) 61.9% of the time. During the EC condition, participants reciprocated after their giving was reciprocated 78.5% of the time, but only 53.5% when their giving was not reciprocated. The binary logistic model showed a significant effect of Reciprocity behavior in the Baseline condition ($p < 0.001$). See Table 2.

The Personal Interest score (DIT-2) correlated positively with reciprocity behavior in the Baseline condition. Within the Test condition, altruistic behavior correlated positively with the Personal Interest score ($R^2 = 0.225$, $p = 0.040$) and negatively with the Postconventional score of the DIT-2. However, these trends did not survive false discovery rate (FDR) correction for multiple comparisons [32].

Regarding credibility of the deception of using computer generated investor behavior, all subjects reported believing that they have played with an assortment of local university students. To illustrate the point, the 10 individuals whose behavior was always to give back regardless or the other players' behavior explained their behavior by several altruistic explanations variants of these two themes: "they probably really needed the money", "if they were already upset I didn't want to be mean to them".

Table 2
Individual traits, and Trust Game performance of subjects included in imaging analysis.

Quantitative assessments	Mean ± standard deviation
Neuroticism (NEO-FFI[a])	19.2 ± 9.1
Extraversion (NEO-FFI)	31.3 ± 6.2
Impulsivity (BIS−11[b])	59.9 ± 11.8
Childhood trauma (CTQ[c])	37.8 ± 13.4
Defining Issues Test-2	
Personal Interest	22.6 ± 12.9
Maintain Norms	29.6 ± 12.9
Postconventional	43.9 ± 17.2
Trust Game performance	
Reciprocity	61.9 ± 22.1
Altruistic behavior (giving back after not receiving)	58.5 ± 31.4
Selfish behavior (not giving back after receiving)	21.5 ± 23.2

[a] Neuroticism-Extraversion-Openness-Five-Factor Inventory (NEO-FFI).
[b] Barratt Impulsiveness Scale-11 (BIS-11).
[c] Childhood Trauma Questionnaire (CTQ).

3.2. Univariate neuroactivation

During the Baseline condition, the Gave $0 contrast was associated with activation of the precentral gyrus, culmen and the superior temporal gyrus.

During the Test condition the contrast Gave $1.5X After Receiving was associated with activation in the inferior temporal gyrus and middle frontal gyrus. The contrast Gave $1.5X After Not Receiving was associated with cuneus activation. Gave $1.5X After Not Receiving versus Gave $0 After Not Receiving was associated with activation in the lingual gyrus (Table 3) and Fig. 2.

3.3. Network neuroactivation

The seven ICA-derived components (IC) associated with hot and/or cold cognition corresponded to four well-identified and spatially independent networks of brain activation: DMN (IC 17), CEN (ICs 9 and 14), LIM (IC 19), and SAL (ICs 2 and 8) (ESM Table 1 and Fig. 2). IC 8 encompassed two networks − a positively contributing saliency network and a negatively contributing sensorimotor network. ICs relating to visual, sensorimotor, and cerebellar systems are provided in Supplementary Fig. 1.

ICA analyses use data from all voxels. These clusters were arbitrarily thresholded at cluster > 20 and $p < 0.005$ the aid visualization of the networks (Fig. 3).

3.4. Task-related network activity

During the Baseline condition, the SAL (IC 18) [$t(18) = 2.46$, $p < 0.024$] was activated during **reciprocity** choice (Gave $1.5X versus Gave $0), which survived FDR correction ($p < 0.05$). During the Test condition, greater engagement of the SAL network (IC 2) was also correlated with **reciprocity** behavior (Gave $1.5X After Receiving) [$t(18) = 2.34, p < 0.039$], however, it did not survive FDR correction.

3.5. Regression with individual behavior and traits

No significant correlations between brain activity and individual behavior or traits were observed.

4. Discussion

The tendency to reciprocate is diminished by the unfair act of others not reciprocating in a prior exchange, though some individuals persist in reciprocating following a perceived violation of reciprocity by others. This study sought to identify the neural correlates of such behavior after fair versus unfair exchanges. We have identified a set of neural processing networks (DMN, SAL, CEN, and LIM) functionally associated with reciprocity behavior. Specifically, we report greater SAL network activity during reciprocity behavior.

4.1. Reciprocity

Social preference research has proposed different underlying motives for prosocial behavior including a tit-for-tat preference, inequality aversion, intrinsic altruism (warm-glow altruism), and reputation seeking [34]. Other distal causes for prosocial behavior include guilt or masochism [5]. Our study design in which participants always played the trustee role in the Baseline condition excludes a tit-for-tat strategy, whereas the one–round design precludes a reputation-seeking approach by participants. The exclusion of these motivations was further supported by the voiced understanding and subjective participant response

Table 3
Brain region activity associated with univariate analysis in the Trust Game in healthy volunteers.

Contrast	Region (Brodmann area)	Cluster size (no. of voxels)	Peak voxel *t*-score	Peak voxel MNI [a] coordinates		
				x	y	z
Baseline condition						
Gave $0						
	Precentral gyrus (l 6)	114	−6.6	−10	−34	63
	Cerebellum culmen (l)	56	8.0	−38	−48	−30
	Cerebellum culmen (r)	47	6.8	31	−48	−61
	Superior temporal gyrus (l 38)	47	−10.6	−28	18	−36
Test condition						
Gave $1.5 After Receiving						
	Inferior temporal gyrus (r 20)	94	13.21	52	−20	−26
	Middle frontal gyrus (r 11)	52	9.44	10	49	−16
	Middle frontal gyrus (l 9)	49	−10.6	−48	18	29
Gave $1.5X After Not Receiving						
	Cuneus (l 18)	48	8.1	−14	−96	19
Altruistic choices (Gave $1.5X After Not Receiving − Gave $0 After Not Receiving)						
	Lingual gyrus (l 19)	69	8.1	−34	−47	1

[a] MNI Montreal Neurological Institute.

to these elements of the game.

Reciprocity behavior is thought to reflect the interaction of anterior medial prefrontal cortex (mPFC) and temporoparietal junction (TPJ). Activation of the anterior mPFC is associated with failure to reciprocate, irrespective of perceptions of risk or benefit, or individual differences in social value orientation. In contrast, activity of the TPJ, bilateral insula, and ACC (regions incorporated in our SAL network) is modulated by individual social value orientation and is sensitive to perceptions of risk and benefit [35]. Additionally, reciprocity behavior positively correlates with BOLD activity in the dorsal ACC, bilateral anterior insula, right TPJ, and precuneus [35].

Cooperation during the Prisoner's Dilemma is associated with activation of the striatum, VMPFC, and ACC [9]. In a task of strategic uncertainty, the payoff dominant option or choice to "cooperate" is associated with posterior STS, ACC, and PCC [36]. Additionally, equitable decisions in the Dictator Game are associated with orbitofrontal cortex (OFC) activation, suggestive of an intrinsic subjective value [37]. Developmental data shows a temporal correlation between the emergence of reciprocity and other prosocial correlates with a functional switch from the anterior mPFC to the TPJ during adolescence [38]. Moreover, anatomical and lesion studies highlight the role of the mPFC in prosocial behavior. For instance, regional grey volume in bilateral PFC correlates with the trait of cooperativeness in young women [39]. Also, individuals with VMPFC lesions show lower reciprocity than healthy controls in the Trust Game [40], although these findings might be explained by decreased ability to experience guilt in patients with VMPFC lesions [41]. This decreased ability, in turn, may explain apparent deficits of these patients during moral judgment [42,43].

Even though the PFC and TPJ are clearly involved in reciprocity, such a complex behavioral outcome like reciprocity behavior is likely the product of the combined output of brain networks involved in emotional, cognitive control, theory of mind (ToM; the uniquely human ability to contemplate others' thoughts, desires, and intentions), and salience processing [38].

4.2. Brain circuits

ICA allowed us to identify neural networks underlying cognitive processing of reciprocity, including DMN, CEN, SAL and LIM. The DMN encompasses an integrated system for self-related cognitive activity including autobiographical, self-monitoring and social cognitive functions [27]. Brain regions associated with the DMN include mPFC, medial temporal lobe, precuneus, PCC, and TPJ. The DMN has considerable overlap with two other neural networks. Although not formally recognized as a moral neural network, an array of brain regions has been consistently activated in a variety of experimental paradigms, including passive picture viewing [17], exposure to morally-laden scenarios [15,44], and active decision making [8], has been associated with the mPFC, ACC, PCC, STS, insula, and amygdala. Secondly, a number of studies in social cognition report activation of brain regions implicated in ToM, including the TPJ and the precuneus [45,46]. The TPJ has been noted to participate in the modulation of justice reasoning [15,44,47].

The CEN is equipped to operate on identified relevant stimuli for the individual, directing attention to pertinent stimuli as behavioral choices are weighed against shifting conditions, background homeostatic demands, and context [48]. To achieve this level of response flexibility, prefrontal and posterior parietal cortex must exert control over sensorimotor representations and maintain relevant data in mind until actions are selected.

The neural processes underlying reciprocity seem to involve both emotion and social cognition [35,49]. This involvement is usually reflected in limbic activity, namely striatum, amygdala, and insula. For instance, amygdala activation is associated with intensity of emotion during human reciprocal interaction [50]. Altruistic punishment has been shown to involve the anterior insula [51] and amygdala [52,53] in addition to striatal reward centers [54].

In contrast, our ICA findings associated the SAL with reciprocity choices. Moreover, brain activity in insula and anterior cingulate behavior has been linked with other prosocial behaviors like trust and fairness in meta-analytic studies [28,30,31]. On the other hand, self-centered behavior is associated with insular activation. In an elegant adaptation of the Trust Game [55], ACC and fronto-insular cortex (SAL network) activity predicted future deception and non-reciprocity behavior. Interestingly, insular activation continued to be present until the time when subjects did not actually reciprocate. Further, inequitable choices in the Dictator Game produced anterior insula activation [37]. Because humans weigh social rewards with economic self-interest, the DLPFC and the VMPFC are likely to be crucially involved in the balancing of competing rewards [2,56]. However, insular activation when choosing the monetary reward may reflect distaste for abandoning a natural prosocial impulse. Interestingly, the SAL association with both pro- (reciprocity) and anti-social (selfish behavior) may be explained by its intimate roles in the processing of subjective value which has been hypothesized to follow a U-shape function [29].

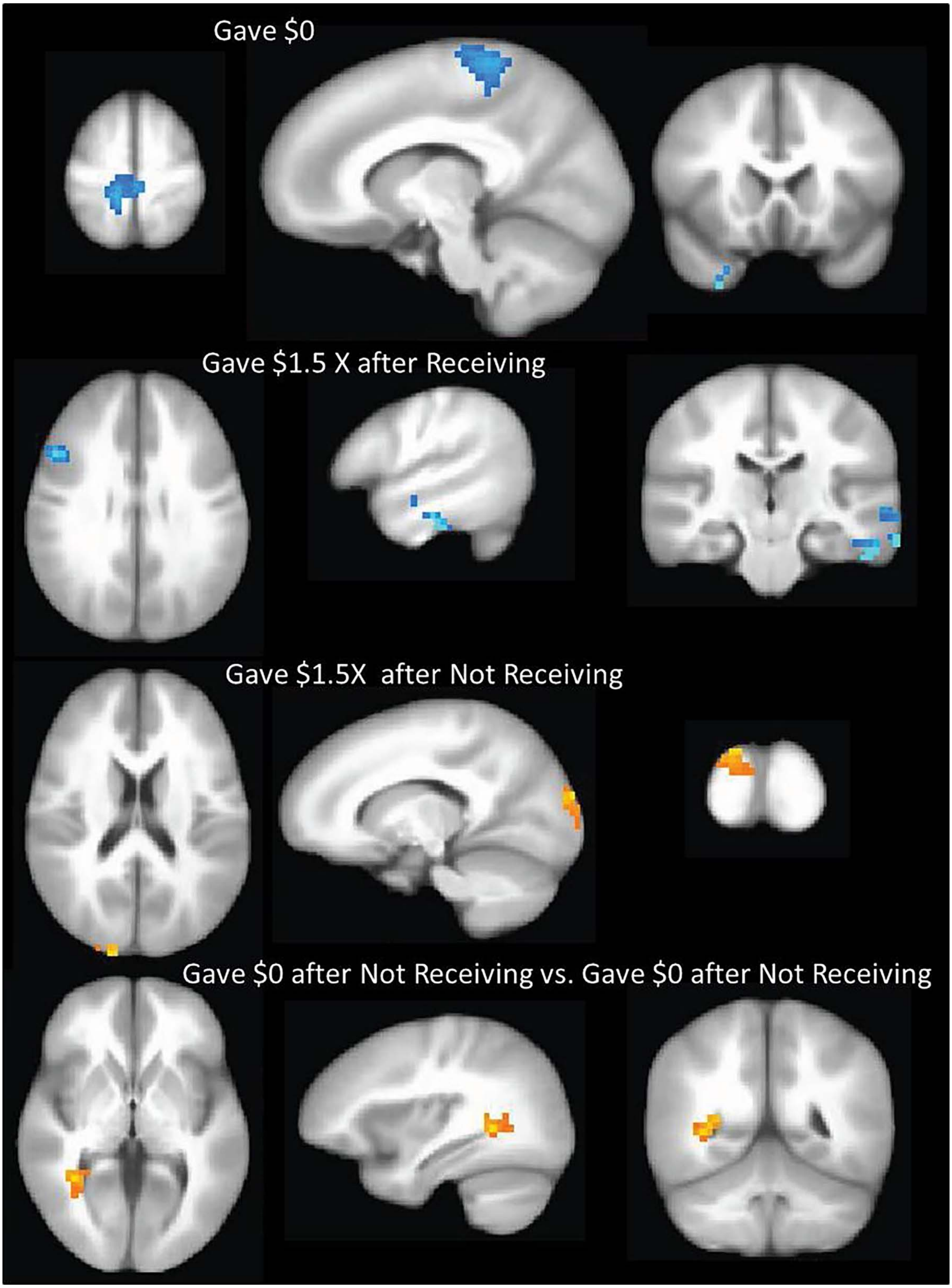


Fig. 2. Brain regions activated during the Baseline condition (Gave $0) and during the Test condition (Gave $1.5X After Receiving, Gave $1.5X After Not Receiving, and Gave $1.5X After Not Receiving versus Gave $0 After Not Receiving).

4.3. Individual traits

The range of Postconventional scores in our sample is comparable to those found in college students and the general population [57]. Even though the correlations between reciprocity and altruistic behavior with moral development measures did not survive multiple comparison correction, it is worth discussing these behaviors in the context of moral reasoning. Moral reasoning refers to the sorting of pros and cons, conflicting values, and duties in order to determine a course of action when confronted by an ethical dilemma [58]. Postconventional scores are associated with prosocial behavior in real life such as community involvement, whistle-blowing and not cheating [59]. The seemingly

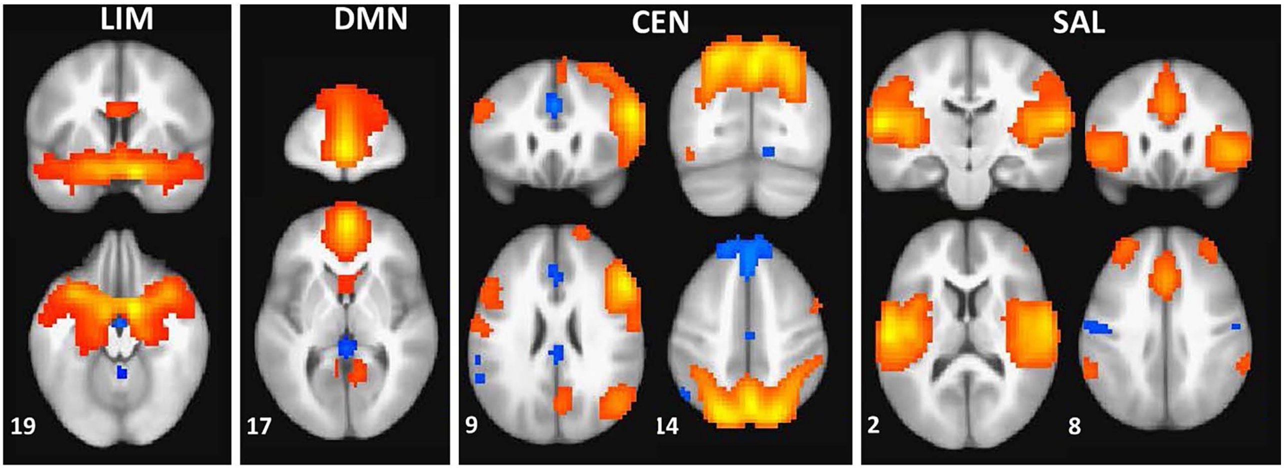


Fig. 3. Brain networks engaged during the Trust Game, as identified by independent component analysis (ICA). The six ICA components corresponded to four brain networks: DMN default-mode network, CEN central executive network, LIM mesolimbic system network, SAL salience network.

counterintuitive positive correlation of reciprocity behavior with the Personal Interest scores highlights the complexity of moral behavior, which depends not only on the level of moral development and moral reasoning capabilities of an individual, but also his/her moral motivation and moral courage that independently lead to a behavioral outcome (whether or not to reciprocate) [58].

4.4. Limitations

Even though the Trust Game we used is a routinely employed behavioral economic tool [60], it is still a simplified model of human social interaction. We did not record income of participants, which has been described as a strong predictor of charitable contributions [61]. We did not measure oxytocin or sexual hormones levels, which have been shown to strongly modulate prosocial behavior [62,63]. Lastly, given the limited sample size we were not able to stratify by gender.

5. Conclusions

Our findings of an association of distinct neural networks (DMN, SAL, CEN and LIM) with reciprocity behavior during the Trust Game add to a growing body of evidence linking involvement of a recurrent set of brain regions, i.e precuneus, superior temporal cortex, insula, medial prefrontal cortex and anterior cingulate, with complex moral emotions and behavior. Even though it may be premature to name these regions a moral network, their consistent involvement in moral-related experiments supports at least a role for engaging of emotional, self-referential and planning processes.

Funding

This work was supported by the Ethics and Community Program from the Arsht Foundation.

Acknowledgments

Editing assistance was provided by the Office of Grants and Scientific Publications, University of Arkansas for Medical Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbr.2017.05.030.

References

[1] R. Axelrod, W.D. Hamilton, The evolution of cooperation, Science 211 (4489) (1981) 1390–1396.
[2] E. Fehr, B. Rockenbach, Human altruism: economic, neural, and evolutionary perspectives, Curr. Opin. Neurobiol. 14 (6) (2004) 784–790.
[3] S.F. Brosnan, L. Salwiczek, R. Bshary, The interplay of cognition and cooperation, Philos. Trans. R Soc. Lond. B Biol. Sci. 365 (1553) (2010) 2699–2710.
[4] J.C. Cox, How to identify trust and reciprocity, Games Econ. Behav. 46 (2) (2004) 260–281.
[5] B.J. Seelig, L.S. Rosof, Normal and pathological altruism, J. Am. Psychoanal. Assoc. 49 (3) (2010) 933–959.
[6] A. Lotem, M.A. Fishman, L. Stone, From reciprocity to unconditional altruism through signalling benefits, Proc. Biol. Sci. 270 (1511) (2003) 199–205.
[7] M. Mohtashemi, L. Mui, Evolution of indirect reciprocity by social information: the role of trust and reputation in evolution of altruism, J. Theor. Biol. 223 (4) (2003) 523–531.
[8] J.D. Greene, R.B. Sommerville, L.E. Nystrom, J.M. Darley, J.D. Cohen, An fMRI investigation of emotional engagement in moral judgment, Science 293 (5537) (2001) 2105–2108.
[9] J. Rilling, D. Gutman, T. Zeh, G. Pagnoni, G. Berns, C. Kilts, A neural basis for social cooperation, Neuron 35 (2) (2002) 395–405.
[10] J. Katsyri, R. Hari, N. Ravaja, L. Nummenmaa, The opponent matters: elevated FMRI reward responses to winning against a human versus a computer opponent during interactive video game playing, Cereb. Cortex 23 (12) (2013) 2829–2839.
[11] K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, A functional imaging study of cooperation in two-person reciprocal exchange, Proc. Natl. Acad. Sci. U. S. A. 98 (20) (2001) 11832–11835.
[12] F. Krueger, K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke, J. Grafman, Neural correlates of trust, Proc. Natl. Acad. Sci. U. S. A. 104 (50) (2007) 20084–20089.
[13] B. King-Casas, D. Tomlin, C. Anen, C.F. Camerer, S.R. Quartz, P.R. Montague, Getting to know you: reputation and trust in a two-person economic exchange, Science 308 (5718) (2005) 78–83.
[14] D. Tomlin, M.A. Kayali, B. King-Casas, C. Anen, C.F. Camerer, S.R. Quartz, P.R. Montague, Agent-specific responses in the cingulate cortex during economic exchanges, Science 312 (5776) (2006) 1047–1050.
[15] R. Cáceda, G.A. James, T.D. Ely, J.R. Snarey, C.D. Kilts, Mode of effective connectivity within a putative neural network differentiates moral cognitions related to care and justice ethics, PLoS One 6 (2) (2011) e14730.
[16] J.D. Greene, L.E. Nystrom, A.D. Engell, J.M. Darley, J.D. Cohen, The neural bases of cognitive conflict and control in moral judgment, Neuron 44 (2) (2004) 389–400.
[17] J. Moll, R. de Oliveira-Souza, P.J. Eslinger, I.E. Bramati, J. Mourao-Miranda, P.A. Andreiuolo, L. Pessoa, The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions, J. Neurosci. 22 (7) (2002) 2730–2736.
[18] P.T. Costa, R.R. McCrae, Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-factor Inventory (NEO-FFI) Professional Manual, Psychological Assessment Resources, Odessa, FL, 1992.
[19] J.R. Rest, D. Narvaez, S.J. Thoma, M.J. Bebeau, DIT2: Devising and testing a revised instrument of moral judgment, J. Educ. Psychol. 91 (4) (1999) 644–659.
[20] J.H. Patton, M.S. Stanford, E.S. Barratt, Factor structure of the Barratt impulsiveness scale, J. Clin. Psychol. 51 (6) (1995) 768–774.
[21] R. Caceda, G.A. James, D.A. Gutman, C.D. Kilts, Organization of intrinsic functional brain connectivity predicts decisions to reciprocate social behavior, Behav. Brain Res. 292 (2015) 478–483.
[22] R. Cáceda, T. Moskovciak, S. Prendes-Alvarez, J. Wojas, A. Engel, S.H. Wilker, J.L. Gamboa, Z.N. Stowe, Gender-specific effects of depression and suicidal ideation in prosocial behaviors, PLoS One 9 (9) (2014) e108733.

[23] R.W. Cox, AFNI. software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. 29 (3) (1996) 162–173.
[24] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith, Fsl, Neuroimage 62 (2) (2012) 782–790.
[25] V.D. Calhoun, T. Adali, G.D. Pearlson, J.J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, Hum. Brain Mapp. 14 (3) (2001) 140–151.
[26] K.L. Ray, D.R. McKay, P.M. Fox, M.C. Riedel, A.M. Uecker, C.F. Beckmann, S.M. Smith, P.T. Fox, A.R. Laird, ICA model order selection of task co-activation networks, Frontiers in neuroscience 7 (2013) 237.
[27] R.N. Spreng, R.A. Mar, A.S. Kim, The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis, J. Cogn. Neurosci. 21 (3) (2009) 489–510.
[28] C. Feng, Y.J. Luo, F. Krueger, Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis, Hum. Brain Mapp. 36 (2) (2015) 591–602.
[29] O. Bartra, J.T. McGuire, J.W. Kable, The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value, Neuroimage 76 (2013) 412–427.
[30] G. Bellucci, S.V. Chernyak, K. Goodyear, S.B. Eickhoff, F. Krueger, Neural signatures of trust in reciprocity: a coordinate-based meta-analysis, Hum. Brain Mapp. 38 (3) (2017) 1233–1248.
[31] A.S. Gabay, J. Radua, M.J. Kempton, M.A. Mehta, The Ultimatum Game and the brain: a meta-analysis of neuroimaging studies, Neurosci. Biobehav. Rev. 47 (2014) 549–558.
[32] C.R. Genovese, N.A. Lazar, T. Nichols, Thresholding of statistical maps in functional neuroimaging using the false discovery rate, Neuroimage 15 (4) (2002) 870–878.
[33] T.D. Wager, M.C. Keller, S.C. Lacey, J. Jonides, Increased sensitivity in neuroimaging analyses using robust regression, Neuroimage 26 (1) (2005) 99–113.
[34] E. Fehr, C.F. Camerer, Social neuroeconomics: the neural circuitry of social preferences, Trends Cogn. Sci. 11 (10) (2007) 419–427.
[35] W. van den Bos, E. van Dijk, M. Westenberg, S.A. Rombouts, E.A. Crone, What motivates repayment? Neural correlates of reciprocity in the Trust Game, Soc. Cogn. Affect. Neurosci. 4 (3) (2009) 294–304.
[36] W.G. Ekins, R. Cáceda, C.M. Capra, G.S. Berns, You cannot gamble on others: dissociable systems forstrategic uncertainty and risk in the brain, J. Econ. Behav. Organ. 94 (2013) 222–233.
[37] J. Zaki, J.P. Mitchell, Equitable decision making is associated with neural markers of intrinsic value, Proc. Natl. Acad. Sci. U. S. A. 108 (49) (2011) 19761–19766.
[38] W. van den Bos, E. van Dijk, M. Westenberg, Rombouts S.A.R.B, E.A. Crone, Changing brains, changing perspectives the neurocognitive development of reciprocity, Psychol. Sci. 22 (1) (2011) 60–70.
[39] H. Yamasue, O. Abe, M. Suga, H. Yamada, M.A. Rogers, S. Aoki, N. Kato, K. Kasai, Sex-linked neuroanatomical basis of human altruistic cooperativeness, Cereb. Cortex 18 (10) (2008) 2331–2340.
[40] G. Moretto, M. Sellitto, G. di Pellegrino, Investment and repayment in a trust game after ventromedial prefrontal damage, Front. Hum. Neurosci. 7 (2013).
[41] I. Krajbich, R. Adolphs, D. Tranel, N.L. Denburg, C.F. Camerer, Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex, J. Neurosci. 29 (7) (2009) 2188–2192.
[42] E. Ciaramelli, M. Muccioli, E. Ladavas, G. di Pellegrino, Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex, Soc. Cogn. Affect. Neurosci. 2 (2) (2007) 84–92.
[43] M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, A. Damasio, Damage to the prefrontal cortex increases utilitarian moral judgements, Nature 446 (7138) (2007) 908–911.
[44] D. Robertson, J. Snarey, O. Ousley, K. Harenski, F. DuBois Bowman, R. Gilkey, C. Kilts, The neural processing of moral sensitivity to issues of justice and care, Neuropsychologia 45 (4) (2007) 755–766.
[45] S.J. Carrington, A.J. Bailey, Are there theory of mind regions in the brain? a review of the neuroimaging literature, Hum. Brain Mapp. 30 (8) (2009) 2313–2335.
[46] F. Van Overwalle, K. Baetens, Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis, Neuroimage 48 (3) (2009) 564–584.
[47] K.J. Yoder, J. Decety, The good, the bad, and the just: justice sensitivity predicts neural response during moral evaluation of actions performed by others, J. Neurosci. 34 (12) (2014) 4161–4166.
[48] W.W. Seeley, V. Menon, A.F. Schatzberg, J. Keller, G.H. Glover, H. Kenna, A.L. Reiss, M.D. Greicius, Dissociable intrinsic connectivity networks for salience processing and executive control, J. Neurosci. 27 (9) (2007) 2349–2356.
[49] A. Strobel, J. Zimmermann, A. Schmitz, M. Reuter, S. Lis, S. Windmann, P. Kirsch, Beyond revenge: neural and genetic bases of altruistic punishment, Neuroimage 54 (1) (2011) 671–680.
[50] S. Sakaiya, Y. Shiraito, J. Kato, H. Ide, K. Okada, K. Takano, K. Kansaku, Neural correlate of human reciprocity in social interactions, Front. Neurosci. 7 (2013) 239.
[51] A.G. Sanfey, J.K. Rilling, J.A. Aronson, L.E. Nystrom, J.D. Cohen, The neural basis of economic decision-making in the Ultimatum Game.[comment], Science 300 (5626) (2003) 1755–1758.
[52] K. Gospic, E., Mohlin, P., Fransson, P., Petrovic, M., Johannesson, M. Ingvar, Limbic justice–amygdala involvement in immediate rejection in the Ultimatum Game, PLoS Biol 9(5) (2011) e1001054.
[53] A.P. Association,, Diagnostic and Statistical Manual of Mental Disorders, 5th ed., American Psychiatric Publishing, Arlington, VA, 2013.
[54] D.J.F. de Quervain, U. Fischbacher, V. Treyer, M. Schelthammer, U. Schnyder, A. Buck, E. Fehr, The neural basis of altruistic punishment, Science 305 (5688) (2004) 1254–1258.
[55] T. Baumgartner, U. Fischbacher, A. Feierabend, K. Lutz, E. Fehr, The neural circuitry of a broken promise, Neuron 64 (5) (2009) 756–770.
[56] S.S. Ho, R.D. Gonzalez, J.L. Abelson, I. Liberzon, Neurocircuits underlying cognition-emotion interaction in a social decision making context, Neuroimage 63 (2) (2012) 843–857.
[57] L.L. Swisher, Moral reasoning among physical therapists: results of the defining issues test, Physiother. Res. Int. 15 (2) (2010) 69–79.
[58] J. Rest, D. Narvaez, M.J. Bebeau, S.J. Thoma, A shorter overview focused on DIT research is: a neo-Kohlbergian approach: the DIT and Schema Theory, Educ. Psychol. Rev. 11 (4) (1999) 291–324.
[59] J.R. Rest, Postconventional moral thinking: a neo-Kohlbergian approach, L. erlbaum associates, mahwah, NJ (1999).
[60] N.D. Johnson, A.A. Mislin, Trust games: a meta-analysis, J. Econ. Psychol. 32 (5) (2011) 865–889.
[61] S.T. Yen, An econometric analysis of household donations in the USA, Appl. Econ. Lett. 9 (13) (2002) 837–841.
[62] J.K. Rilling, A.C. DeMarco, P.D. Hackett, R. Thompson, B. Ditzen, R. Patel, G. Pagnoni, Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men, Psychoneuroendocrinology 37 (4) (2012) 447–461.
[63] P.J. Zak, K. Borja, W.T. Matzner, R. Kurzban, The neuroeconomics of distrust: sex differences in behavior and physiology, Am. Econ. Rev. 95 (2) (2005) 360–363.