RESEARCH ARTICLE

Connectome-based model predicts individual differences in propensity to trust

Xiaping Lu^{1,2,3†} | Ting Li^{4†} | Zhichao Xia³ | Ruida Zhu³ | Li Wang⁴ | Yue-Jia Luo^{1,5,6} | Chunliang Feng^{1,7} | Frank Krueger^{8,9}

¹Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, China

²Brain, Mind & Markets Laboratory, Department of Finance, The University of Melbourne, Melbourne, Victoria, Australia

³State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

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

⁵Center for Emotion and Brain, Shenzhen Institute of Neuroscience, Shenzhen, China

⁶Medical School, Kunming University of Science and Technology, Kunming, China

⁷College of Information Science and Technology, Beijing Normal University, Beijing, China

⁸School of Systems Biology, George Mason University, Fairfax, Virginia

⁹Department of Psychology, University of Mannheim, Mannheim, Germany

Correspondence

Yue-Jia Luo, Ph.D., Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen 518060, China.

Email: luoyj@szu.edu.cn

and

Chunliang Feng, Ph.D., Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, 518060, China.

Email: chunliang.feng@bnu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 31530031; China Postdoctoral Science Foundation, Grant/Award Number: 2017M610055; National Postdoctoral Program for Innovative Talents, Grant/Award Number: BX201600019

Abstract

Trust constitutes a fundamental basis of human society and plays a pivotal role in almost every aspect of human relationships. Although enormous interest exists in determining the neuropsychological underpinnings of a person's propensity to trust utilizing task-based fMRI; however, little progress has been made in predicting its variations by task-free fMRI based on whole-brain resting-state functional connectivity (RSFC). Here, we combined a one-shot trust game with a connectome-based predictive modeling approach to predict propensity to trust from whole-brain RSFC. We demonstrated that individual variations in the propensity to trust were primarily predicted by RSFC rooted in the functional integration of distributed key nodes—caudate, amygdala, lateral prefrontal cortex, temporal-parietal junction, and the temporal pole—which are part of domain-general large-scale networks essential for the motivational, affective, and cognitive aspects of trust. We showed, further, that the identified brain-behavior associations were only evident for trust but not altruistic preferences and that propensity to trust (and its underlying neural underpinnings) were modulated according to the extent to which a person emphasizes general social preferences (i.e., horizontal collectivism) rather than general risk preferences (i.e., trait impulsiveness). In conclusion, the employed data-driven approach enables to predict propensity to trust from RSFC and highlights its potential use as an objective neuromarker of trust impairment in mental disorders.

KEYWORDS

connectome-based predictive modeling, individual difference, resting-state functional connectivity, social decision making, trust

1 | INTRODUCTION

Trust pervades human social life and plays a critical role in numerous social interactions, from facilitating family relations and friendship to

promoting and maintaining social exchange among strangers. Representing a social dilemma, trust entails a trustor's willingness to be vulnerable to the risk of betrayal based on the expectations that the action of a trustee will produce some anticipated reward because of reciprocity in the future (Seppänen, Blomqvist, & Sundqvist, 2007). In a laboratory setting, a person's propensity to trust can be measured

[†]Xiaping Lu and Ting Li contributed equally to the current work.

with the one-shot version of the trust game (TG; Berg, Dickhaut, & McCabe, 1995; Camerer & Weigelt, 1988). In the TG, trustors decide how much to share from their monetary endowment with trustees. The shared money is then tripled in value by the experimenter and sent to the trustees. The trustees decide then how much to return to the trustors but can also decide to return nothing. The amount of money shared by trustors measures trust propensity since trustors make themselves vulnerable to the potential betrayal of trustees (Fehr, 2009).

People exhibit some extent of trust by sharing a portion of money with strangers (Johnson & Mislin, 2011), but they are heterogeneous regarding their propensity to trust (Cesarini et al., 2008; Krueger et al., 2012). Individual variations in propensity to trust have been related to many factors, including genetic polymorphisms (Krueger et al., 2012; Nishina, Takagishi, Inoue-Murayama, Takahashi, & Yamagishi, 2015), plasma oxytocin levels (Zhong et al., 2012), and personality attributes (Alarcon et al., 2018; Unoka, Seres, Áspán, Bódi, & Kéri, 2009), such as individual-level collectivism and individualism (Shin & Park, 2005; Van Lange, Rockenbach, & Yamagishi, 2017; Zeffane, 2017). Here, we examined whether individual variations in the trust propensity can be predicted from intrinsic brain imaging measures.

The past decades have witnessed numerous neuroimaging studies on the trust (Delgado, Frank, & Phelps, 2005; King-Casas et al., 2005; Krueger et al., 2007; McCabe, Houser, Ryan, Smith, & Trouard, 2001). In light of previous findings, a neuropsychoeconomic model (Krueger & Meyer-Lindenberg, 2018) suggest that trust arises through the interplay of psychological components (i.e., motivation, affect, cognition) that engage key brain regions anchored in domain-general large-scale brain networks: reward network (RWN: e.g., *striatum*), salience network (SAN: e.g., amygdala; anterior insula, AI); central-executive network (CEN: e.g., lateral prefrontal cortex, lateral PFC); and default-mode network (DMN: e.g., temporoparietal junctions, TPJ; and dorsomedial prefrontal cortex, dmPFC). In particular, the anticipation of reward (motivation, RWN) contrasted with the aversive feeling for risk of betrayal (affect, SAN) creates uncertainty for the vulnerability of trusting another person. To remove uncertainty, an adoption of a context-based strategy (cognitive control, CEN) or evaluation of relationship-based trustworthiness (social cognition, DMN) can be employed, thereby, transforming risk of betrayal to positive expectations of reciprocity.

The neuropsychological underpinnings of trust have been studied dominantly by utilizing task-based fMRI, but it remains obscure whether trust propensity can be predicted by task-free fMRI based on resting-state functional connectivity (RSFC). As a task-independent measure, RSFC is free from confounds associated with ongoing task demand and different experimental designs across studies (Kable & Levy, 2015; Nash, Gianotti, & Knoch, 2015; Nash & Knoch, 2016). Moreover, fMRI-RSFC has emerged as a widely-used network-level approach to significantly advance our understanding of individual variations in cognitive functions, personality traits, and behaviors (Feng, Yuan, et al., 2018; Feng, Zhu, et al., 2018; Finn et al., 2015; Gianotti, Lobmaier, Calluso, Dahinden, & Knoch, 2017; Gianotti, Nash, Baumgartner, Dahinden, & Knoch, 2018; Hsu,

Rosenberg, Scheinost, Constable, & Chun, 2018; Jung, Lee, Lerman, & Kable, 2018; Rosenberg et al., 2016; Wang et al., 2017).

A multivariate machine learning approach is particularly suitable for examining RSFC underlying complex and multifaceted constructs such as propensity of trust. For example, previous evidence has that a multivariate predictive model combining the multi-round TG with electroencephalography (EEG)-based RSFC predicts only initial trust as measured during the first round (i.e., trust propensity) but not trust as measured across multiple rounds (i.e., trust dynamics; Hahn et al., 2014). Further, the machine learning approach allows for the prediction of unseen participants, offering information at the individual level rather than group level (Cui, Su, Li, Shu, & Gong, 2018; Cui, Xia, Su, Shu, & Gong, 2016; Dubois & Adolphs, 2016; Feng et al., 2017; Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015; Shen et al., 2017; Yarkoni & Westfall, 2017). A machine learning approach typically implements cross-validation procedures to estimate the model with training samples and to test the performance of the model with independent samples (i.e., test samples).

In this study, we implemented a connectome-based predictive modeling approach to predict individual variations of trust propensity from whole-brain RSFC. Based on previous evidence, we hypothesized that individual differences in trust propensity would be predicted by functional interplay between a wide array of functional connectivity across distributed networks, particularly those implicated in valuation (RWN: e.g., caudate), aversive emotional processing (SAN: e.g., amygdala, AI), context-based strategy (CEN: e.g., lateral PFC), and relationship-based trustworthiness (DMN: e.g., TPJ, dmPFC).

2 | MATERIALS AND METHODS

2.1 | Participants

The whole sample included 168 healthy right-handed college students (109 males; 21.99 ± 2.38 years old, range: 18–30 years old) without history of neurological or psychiatric disorder. Among them, 89 participants (46 males) played economic games (i.e., measuring trust propensity), filled survey measures (see below), and completed a neuroimaging scan. The other participants ($n = 79$, 63 males) only filled survey measures and completed a neuroimaging scan. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Beijing Normal University. Written informed consents were obtained from all participants.

2.2 | Economic exchange games

Participants played multiple rounds ($n = 12$) of the one-shot TG (Berg et al., 1995; Camerer & Weigelt, 1988) and one-shot dictator game (DG; Kahneman, Knetsch, & Thaler, 1986) with different putative anonymous partners without feedback between rounds. Before the games, participants were given written instructions on the payoff and rules for TG and DG. Afterwards, participants answered several questions designed to assess their understanding of the TG and DG.

In the TG, participants acted in the role of investor (i.e., trustors) and started with an endowment of 9 monetary units (MUs). They

needed to decide whether to trust or not by passing any portion of the endowment to the trustees and keep the remainder of the endowment. The shared money would be tripled in value and passed to the anonymous partners, who would putatively participate in the experiment later as trustees to decide how much to return of the received money. The amount of money shared by trustors measured trust propensity since trustors made themselves vulnerable to the potential betrayal of trustees (Fehr, 2009).

In the DG, participants acted as dictators and decided how to split a sum of money (12 MUs) between themselves and the other player as a passive recipient. Participants' behaviors in the DG reflected generosity or altruistic preferences (Benenson, Pascoe, & Radmore, 2007; Zak, Stanton, & Ahmadi, 2007).

We employed participants' initial behaviors (i.e., the first round) in the TG and DG as their measures of trust and altruistic preferences, respectively (Note that we also report the results derived from the behaviors averaging across all rounds for sake of completeness). This is based on Hahn et al. (2014)'s study, in which they found that resting-state brain-electrical connectivity only predicted individual variations in trust measured during the first round (i.e., trust propensity) but not trust measured across multiple rounds (i.e., trust dynamics).

Participants came to the lab only once to complete the experiment—playing first the DG first (as the baseline) followed by the TG. They were informed that they would be paid a week later after putative trustees had made their decisions. Unbeknownst to the participants, no other players were recruited for this experiment. To encourage real decisions, however, it was emphasized that the MUs they earned in the game would be converted into their final monetary payoffs. However, participants did not know the exact exchange rate from MUs to monetary payouts, and each participant was paid a fixed fee (50 CNY). Before leaving the laboratory, participants completed a debriefing questionnaire designed to examine their beliefs about the experimental setup, and none of the participants expressed doubts.

2.3 | Survey measures

Participants were asked to complete the (a) horizontal and vertical individualism–collectivism scale (Singelis, Triandis, Bhawuk, & Gelfand, 1995), measuring subdimensions of vertical individualism (i.e., conception of an autonomous individual and acceptance of inequality), vertical collectivism (i.e., perceiving the self as a part of a collective and accepting inequalities within the collective), horizontal collectivism (i.e., perceiving the self as a part of the collective, but seeing all members of the collective as the same) and horizontal individualism (i.e., conception of an autonomous individual and emphasis on equality) and (b) a brief version Barratt Impulsiveness Scale-11 (BIS Brief; Patton, Stanford, & Barratt, 1995; Steinberg, Sharp, Stanford, & Tharp, 2013), representing an unidimensional trait impulsiveness measure (Steinberg et al., 2013).

2.4 | Image acquisition

Images were acquired with a Siemens Trio 3-Tesla scanner at the Beijing Normal University Imaging Center for Brain Research, Beijing, China. All participants first completed a 5-min resting-state fMRI

scanning, during which they were instructed to close their eyes, keep still, remain awake, and not to think about anything systematically. The resting state scanning consisted of 150 contiguous echo-planar imaging volumes using the following parameters: axial slices, 33; slice thickness, 3.5 mm; gap, 0.7 mm; TR, 2,000 ms; TE, 30 ms; flip angle, 90°; voxel size, $3.5 \times 3.5 \times 3.5 \text{ mm}^3$; and FOV, $244 \times 244 \text{ mm}^2$. In addition, high-resolution structural images were acquired through a 3D sagittal T1-weighted magnetization-prepared rapid acquisition with gradient-echo sequence, using the following parameters: sagittal slices, 144; TR, 2,530 ms; TE, 3.39 ms; slice thickness, 1.33 mm; voxel size, $1 \times 1 \times 1.33 \text{ mm}^3$; flip angle, 7°; inversion time, 1,100 ms; and FOV, $256 \times 256 \text{ mm}^2$.

2.5 | Image preprocessing

Neuroimaging data analyses were performed with the DPABI software plug-in package (<http://rfmri.org/dpabi>; Yan, Wang, Zuo, & Zang, 2016) based on SPM (<http://www.fil.ion.ucl.ac.uk/spm>). The first 10 volumes of the functional images were discarded for signal equilibrium and participants' adaptation to scanning noise. The images were then realigned for head movement correction. Thirteen participants (9 males) were excluded from further analysis under the criteria of head motion exceeding 2.5 mm maximum translation, 2.5° rotation or mean frame-wise displacement exceeding 0.2 mm throughout scans (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012; Yan et al., 2013). To normalize functional images, participants' structural brain images were first co-registered to their mean functional images and were subsequently segmented. The parameters derived from segmentation were used to normalize each participant's functional images into the standard Montreal Neurological Institute space (MNI template, resampling voxel size was $3 \times 3 \times 3 \text{ mm}^3$). Afterwards, the linear trends of time courses were removed, and a band-pass filtering (0.01–0.1 Hz) was applied to the time series of each voxel to reduce the effect of low-frequency drifts and high-frequency physiological noise (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Zuo et al., 2010). Subsequently, the images were spatially smoothed using a Gaussian filter to decrease spatial noise (with a kernel of full width at half maximum 4 mm). Finally, common nuisance variables were regressed out, including the white matter signal, the cerebrospinal fluid signal (Fox et al., 2005; Snyder & Raichle, 2012), and 24 movement regressors including autoregressive models of motion incorporating 6 head motion parameters, 6 head motion parameters one time point before, and the 12 corresponding squared items (Friston, Williams, Howard, Frackowiak, & Turner, 1996).

2.6 | RSFC feature extraction

A 268-node functional brain atlas spanning the whole brain was used (Shen, Papademetris, & Constable, 2010; Shen, Tokoglu, Papademetris, & Constable, 2013), which was derived from a graph-theory based parcellation algorithm that maximized the similarity of the voxel-wise time series within each node (Shen et al., 2013).

For each participant, a time course was computed for each node by averaging the BOLD signal for all the voxels within the node at each time point. Second, network edges were defined as functional

connectivity between each pair of nodes, calculated as the Pearson correlation coefficient between the time courses of each pair of nodes. Fisher's *r*-to-*z* transformation was then implemented to improve the normality of the correlation coefficients, resulting in a 268×268 symmetric connectivity matrix that represented the set of edges/connections in each participant's resting-state connectivity profile. These edges/connections (i.e., connectivity strength) measured as Fisher-transformed correlations were used as features in the predictive models described below.

2.7 | Multivariate relevance vector regression analysis

The relationship between trust propensity and functional connectome was examined using multivariate relevance vector regression (RVR) as implemented in P_{Ro}NTo (<http://www.mlnl.cs.ucl.ac.uk/pronto/>) and in-house scripts running under MATLAB (Mathworks, 2016 release). This algorithm was employed because of the following reasons: (a) our recent study showed that this method works well with high-dimensional features in the prediction analysis (Feng, Cui, Cheng, Xu, & Gu, 2018); (b) this algorithm is relatively simple to implement since there are no hyperparameters to be optimized; and (c) previous studies have shown that RVR outperforms other predictive models (Chu, Ni, Tan, Saunders, & Ashburner, 2011; Franke, Ziegler, Klöppel, Gaser, & Initiative, 2010; Wang, Fan, Bhatt, & Davatzikos, 2010). Notably, we did not perform other prediction algorithms nor compared them with the RVR model, since this issue is beyond the scope of the current work and has been addressed in a recent study (Cui & Gong, 2018).

RVR is a sparse kernel learning multivariate regression method set in a fully probabilistic Bayesian framework (Tipping, 2001). In this framework, a zero-mean Gaussian prior is introduced over the model weights and is governed by a set of hyper-parameters, one for each weight. The most probable values for these hyper-parameters are then iteratively estimated from the training data, with sparseness achieved because of posterior distributions of many of the weights peaking sharply around zero. Those training vectors associated with nonzero weights are referred to as "relevance" vectors. The optimized posterior distribution of the weights can then be used to predict the target value (e.g., trusting preferences) for a previously unseen feature vector, by computing the predictive distribution (Tipping, 2001). In particular, the equation for the RVR model can be represented as below (see also Cui & Gong, 2018; Tipping, 2001):

$$f(x_i) = \sum_{j=1}^p \beta_j x_{i,j} + \beta_0 = \sum_{j=1}^p \left(\sum_{s=1}^l \alpha_s x_{s,j} \right) x_{i,j} + \beta_0 \\ = \sum_{s=1}^l \alpha_s (x_i * x_s) + \beta_0,$$

where x_i is a high-dimensional feature vector ($x_{i,1}, \dots, x_{i,p}$), p is the number of features, β_j is the regression coefficient of the j th feature, l is the quantity of "relevant vectors" (defined below), α_s is the weight for each data point.

Thus, RVR calculates the weight (i.e., the α_s in the formula) for the data points. As RVR is a sparse model in the sample space, most weight would be zero and the remaining samples with nonzero weight were used to fit the model. These samples were called "relevance

vectors." Then, the weight of each feature (i.e., the β_j in the formula) can be calculated using the following formula:

$$\beta_j = \sum_{s=1}^l \alpha_s x_{s,j}.$$

A leave-one-out cross-validation (LOOCV) was used to evaluate the out-of-sample prediction performance. $N - 1$ participants (N is the number of participants) were used as the training set, with the remaining individual used as the testing sample. During the training procedure, an RVR prediction model was constructed using the training set. During the testing procedure, the RVR prediction model was used to predict the unseen testing participants' trusting preferences (Gong et al., 2014). The training and testing procedures were repeated N times such that each subject was used once as the testing subject.

The performance of the prediction was assessed with two frequently used statistics (Franke et al., 2010; Gong et al., 2014): (a) the correlation coefficient (i.e., r) between real and predicted trusting behaviors and (b) the mean squared error (MSE). A permutation test was applied to determine whether the obtained metrics were significantly better than those expected by chance. The trust propensity across training samples were permuted without replacement 5,000 times, and each time the above LOOCV prediction procedure was reapplied. The permutation resulted in a distribution of r and MSE values reflecting the null hypothesis that the model did not exceed chance level. The number of times that the permuted value was greater than (or, concerning MSE values, less than) the true value, was then divided by 5,000, providing an estimated P -value for each statistic.

2.8 | Contributing features and corresponding weights

The absolute value of the RVR weight of each feature (edge) quantifies its contribution to the model (Cui & Gong, 2018; Gong et al., 2014). It is noteworthy that RVR calculates weights for samples. As RVR is a sparse model in the sample space, most weight will be zero; remaining samples with nonzero weight were used to fit the model. The regression coefficients of all features were determined as the weighted sum of the feature vector of the nonzero weighted samples (see also Cui & Gong, 2018; Gong et al., 2014). A larger absolute value of weight indicates a greater contribution of the corresponding feature to prediction, in the context of every other features (Cui & Gong, 2018; Erus et al., 2015; Gong et al., 2014). The feature was selected for visualization if the absolute value of its weight was within the top 10% of the absolute weight values (Ecker et al., 2010; Gong et al., 2014). This threshold was applied to eliminate noise components for a better visualization of the most discriminating regions (Ecker et al., 2010; Gong et al., 2014).

To further characterize the neural substrates of the contributing network, the network was first defined as the set of edges with top 10% weight values described above. Afterwards, the 268 nodes were grouped into 10 macroscale brain regions, including the prefrontal lobe (46 nodes), motor lobe (21 nodes), insula lobe (7 nodes), parietal lobe (27 nodes), temporal lobe (39 nodes), occipital lobe (25 nodes), limbic lobe (36 nodes), cerebellum lobe (41 nodes), subcortical lobe (17 nodes), and brainstem lobe (9 nodes; Finn et al., 2015; Rosenberg

et al., 2016). The number of edges with top 10 weight values between each pair of macroscale regions or canonical networks was then calculated. Finally, the importance of individual nodes was measured as the number of their connections as in previous studies (Beaty et al., 2018; Rosenberg et al., 2016) and the connectivity patterns of the top 10 most highly connected nodes were illustrated.

2.9 | Ten-fold validation

A potential disadvantage of the LOOCV is that it maximizes variance of test-set and may not yield stable estimates of predictive performance. Accordingly, a 10-fold cross-validation was applied to validate the prediction results, since this scheme may provide more stable estimates of predictive performance (Varoquaux et al., 2017). All participants were divided into 10 subsets, of which nine were used as the training set, and the remaining one was as the testing set. The training set was scaled and used to train an RVR prediction model, which was then used to predict the scores for the scaled testing data. The scaling of testing data used parameters acquired from training data. This procedure was repeated 10 times so that each subset was used as a testing set once. Finally, the correlation r and MSE between the true and predicted scores were calculated across all participants. Since the full dataset was randomly divided into 10 subsets, performance might have depended on data division. Therefore, the 10-fold cross-validation was repeated 100 times, and the results averaged to produce a final prediction performance. A permutation test was applied 5,000 times to test the significance of the prediction performance.

2.10 | Control analysis

Control analysis was implemented to further examine the significance of predictions of our models, despite potential confounds of altruistic preference, motion, age, and gender. In the control analysis, the association between actual and predicted trust propensity was recomputed after adjusting these confounding variables.

2.11 | The relationship between trust propensity and survey measures

The association between trust propensity and social preferences (i.e., collectivism/individualism) as well as trait impulsiveness was examined at both behavioral and neural levels. At the behavioral level, the correlations between the trust propensity and the four subscales of the horizontal and vertical individualism–collectivism scale (Singelis et al., 1995) as well as BIS Brief (Steinberg et al., 2013) were assessed. Since trust propensity exhibited only reliable correlation with the horizontal collectivism subscale (see Section 3), it was further investigated whether network features contributing to the prediction of trust propensity can predict horizontal collectivism at the neural level. For this type of analysis, features with top 10% weight values selected from the prediction model of trusting preferences were forwarded to the model predicting the horizontal collectivism. That is, another RVR model was trained to predict horizontal collectivism based on the top 10% predictive edges derived from the trust preferences prediction analysis.

2.12 | Prediction of altruistic preferences in the DG

To further examine whether the brain-behavior relationship was specific to the trust-related preferences, similar RVR models were implemented to reveal whether RSFC profiles could also predict altruistic behaviors in the DG. In the first model, all RSFC features were employed; accordingly, this model is the same as the prediction model for the TG, except that prediction target was altruism in the DG. In the second model, features with top 10% weight values selected from the prediction model of trusting preferences were forwarded to the model predicting the altruistic preferences; therefore, this model specifically examined whether trust-related predictive networks were also associated with altruistic preferences.

3 | RESULTS

3.1 | Behavior

Participants demonstrated individual variations in their trust and altruistic preferences as measured with the initial behavior observed in first round of the multi-round one-shot TG and DG, respectively (Table 1).

3.2 | Predicting trusting propensity from RSFC

A LOOCV approach was implemented to examine whether the relevance between connectivity strength (i.e., Fisher-transformed correlations) and trust preferences (i.e., measured as participants' decisions in the first round of the TG) generalized to unseen individuals. RSFC strength predicted trust propensity in unseen individuals ($r = 0.26$, $p = .024$; $MSE = 2.23$, $p = .019$, permutation test), even after adjusting for altruistic preference, motion, age, and gender ($r = 0.27$, $p = .019$; $MSE = 1.89$, $p = .0012$, permutation test, Figure 1). RSFC strength did not predict trust averaged across multiple rounds ($r = 0.1$, $p = .082$; $MSE = 2.15$, $p = .16$, permutation tests, Supporting Information Figure S1).

TABLE 1 Behavioral data

	M	SD	MIN	MAX
TG	3.81	1.54	0	8
DG	4.65	1.88	0	6
HC	38.27	6.54	12	54
HI	42.57	5.67	20	56
VC	35.5	5.83	16	51
VI	36.63	4.81	25	51
BIS brief	18.05	3.68	9	29

Note that participants started with an endowment of 9 MUs for trust game (TG) and 12 MUs for dictator game (DG). Scores for horizontal collectivism (HC), horizontal individualism (HI), vertical collectivism (VC), and vertical individualism (VI) could range from 8 to 56; and scores for brief version Barratt Impulsiveness Scale-11 (BIS brief) from 8 to 32. Abbreviations: M = mean; MAX = maximum; MIN = minimum; MUs = monetary units; SD = standard deviation.

3.3 | Contributing networks in the prediction of trust propensity

The network consisted of features with top 10% weight scores included 3,578 edges. Based on macroscale regions, (a) connections within prefrontal, subcortical, limbic, parietal, and temporal lobes, (b) connections between prefrontal and these lobes, (c) connections between subcortical and temporal and limbic lobes, and (d) connections between limbic and parietal, temporal, and occipital lobes were primary predictors of trust propensity (Figure 2a,b).

Furthermore, the top 10 most highly connected nodes were located in the caudate, amygdala, middle frontal gyrus (MFG), inferior frontal gyrus (IFG), temporoparietal junction (TPJ), temporal pole (TP), cuneus, and cerebellum (Figure 3 and Table 2).

TABLE 2 Ten nodes with the most connections contributing to the prediction model

Node	MNI coordinates (mm)			Lobe	Degree
L caudate	-10.8	23.9	9.6	Subcortical	120
L TPJ	-51.4	-56.3	20.5	Parietal	75
L TP	-34.6	18.6	-32.2	Temporal	73
R MFG	39.7	3.4	34	Prefrontal	70
R cerebellum	39.1	-74.9	-29.7	Cerebellum	65
R cerebellum	33.8	-47.2	-53.9	Cerebellum	65
R caudate	21.2	-36.4	22.6	Subcortical	64
L cuneus	-6	-81.2	12.2	Occipital	63
L amygdala	-22.7	-12.8	-17.4	Limbic	62
L IFG	-32.5	22.1	5.8	Prefrontal	60

IFC = inferior frontal gyrus; L = left; MFG = middle frontal gyrus; R = right; TP = temporal pole; TPJ = temporal-parietal junction. Degree was measured as the number of connected other nodes for each node.

FIGURE 1 Performance of the prediction model assessed by leave-one-out cross-validation. (a) Correlation between actual and predicted trust preferences adjusting for altruistic preferences, motion, gender, and age. (b) Permutation distribution of the correlation coefficient (r) for the prediction analysis. The value obtained using real scores is indicated by the blue dashed line. (c) Consistency between actual and predicted trust preferences adjusting for altruistic preferences, motion, gender, and age. (d) Permutation distribution of the mean squared error. The value obtained using real scores is indicated by the blue dashed line [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Functional connections predicting trust preferences. (a) The connectivity patterns selected by the prediction model, plotted as the number of connections within each macroscale regions. (b) Connections plotted as the number of edges within and between each pair of macroscale regions. L, left; R, right; PFC, prefrontal; Mot, motor; Ins, insula; Par, parietal; Tem, temporal; Occ, occipital; Lim, limbic; Cer, cerebellum; Sub, subcortical; Bsm, brainstem [Color figure can be viewed at wileyonlinelibrary.com]

3.4 | Ten-fold validation

Ten-fold cross-validation was used to validate the performance of the prediction. The correlation coefficient and MSE values remained significant ($r = 0.23$, $p = .037$; $MSE = 2.38$, $p = .028$, permutation test), even after adjusting for altruistic preference, motion, age, and gender ($r = 0.24$, $p = .028$; $MSE = 1.93$, $p = .002$, permutation test, Figure 4).

3.5 | The relationship between trust propensity and survey measures

The correlational analyses revealed that trust propensity was positively correlated only with the horizontal collectivism subscale ($r = 0.32$, $p = .003$) even after adjusting for altruistic preference, age, and gender ($r = 0.39$, $p < .0005$), but not with other subscales ($r_s < 0.1$, $p_s > 0.39$; Figure 5a). However, predicted trust propensity was not correlated with horizontal collectivism subscale ($r = -0.032$, $p = .77$).

To further assess the association between horizontal collectivism and network features contributing to the prediction of trust propensity, it was examined whether these network features can predict horizontal

collectivism. The results revealed that trust-related network features (i.e., the 3,578 edges with top 10% weight values selected from the predictive model of trust propensity) predicted horizontal collectivism ($r = 0.31$, $p = .0018$; $MSE = 41.18$, $p = .0044$, permutation tests) also after adjusting for motion, age, and gender ($r = 0.31$, $p = .0014$; $MSE = 41.56$, $p = .0046$, permutation tests, Figure 5b,c). In line with the behavioral findings, the trust-related predictive network could not reliably predict individual differences for the other subscale scores ($r_s < 0.10$).

No significant correlation was found between trust propensity and trait impulsiveness (unadjusted for covariates: $r = 0.17$, $p = .12$; adjusted for covariates: $r = 0.18$, $p = .11$, Figure 5d), and the trust-related predictive network did not reliably predict trait impulsiveness ($r = -0.05$, Figure 5e,f).

3.6 | Prediction of altruistic preferences

Participants' initial altruistic preferences could not be predicted by either whole-brain functional connectivity profiles (unadjusted for covariates: $r = -0.24$, $p = .92$; $MSE = 5.03$, $p = .96$, permutation tests; adjusted for covariates: $r = -0.24$, $p = .91$; $MSE = 4.93$, $p = .96$, permutation tests,

FIGURE 3 Connectivity patterns of the top ten nodes with the most connections. L, left; R, right; TPJ, temporo-parietal junction; TP, temporal pole; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; PFC, prefrontal; Mot, motor; Ins, insula; Par, parietal; Tem, temporal; Occ, occipital; Lim, limbic; Cer, cerebellum; Sub, subcortical; Bsm, brainstem [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 4 Performance of the prediction model assessed by 10-fold cross-validation. (a) Correlation between actual and predicted trust preferences adjusting for altruistic preferences, motion, gender, and age. (b) Permutation distribution of the correlation coefficient (r) for the prediction analysis. The value obtained using real scores is indicated by the blue dashed line. (c) Consistency between actual and predicted trust preferences adjusting for altruistic preferences, motion, gender, and age. (d) Permutation distribution of the mean squared error. The value obtained using real scores is indicated by the blue dashed line [Color figure can be viewed at wileyonlinelibrary.com]

Supporting Information Figure S2a,b) or trust-related predictive network (unadjusted for covariates: $r = -0.24$, $p = .92$; $MSE = 5.28$, $p = .95$, permutation tests; adjusted for covariates: $r = -0.25$, $p = .92$; $MSE = 5.22$, $p = .96$, permutation tests, Supporting Information Figure S2c,d). Finally, participants' altruistic behaviors averaged across multiple rounds could not be by either whole-brain functional connectivity profiles (Supporting Information Figure S3a,b) or trust-related predictive network (Supporting Information Figure S3c,d).

4 | DISCUSSION

In this study, we applied a connectome-based predictive modeling framework to predict individual variations of trust propensity

(measured with the TG) from intrinsic whole-brain functional connectivity. Overall, our results showed that the functional interplay across multiple neural networks enabled prediction of trust propensity independently of confounding variables such as altruistic preferences, age, gender, and head motion. Inter-individual variations in trust propensity were primarily predicted by RSFC rooted in the functional integration of distributed key nodes within and between domain general large-scale networks—RWN (caudate), SAN (amygdala), CEN (lateral PFC), and DMN (TPJ, TP)—for which neural activity has been previously associated with psychological components (i.e., motivation, affect, and cognition) of trust (Krueger & Meyer-Lindenberg, 2018). Moreover, trust propensity and its underlying neural underpinnings were modulated according to the extent to which a person emphasizes general social preferences (i.e., horizontal collectivism) rather

FIGURE 5 Performance of the prediction model for horizontal collectivism and trait impulsiveness based on trust-related network features, assessed by leave-one-out cross-validation. (a) Correlation between horizontal collectivism and trusting preference. (b) Correlation between actual and predicted horizontal collectivism scores adjusting for motion, gender, and age. (c) Permutation distribution of the correlation coefficient (r) for the prediction analysis of horizontal collectivism. The value obtained using real scores is indicated by the blue dashed line. (d) Correlation between trait impulsiveness and trusting preference. (e) Correlation between actual and predicted trait impulsiveness scores adjusting for motion, gender, and age. (f) Permutation distribution of the correlation coefficient (r) for the prediction analysis of trait impulsiveness. The value obtained using real scores is indicated by the blue dashed line [Color figure can be viewed at wileyonlinelibrary.com]

than general risk preferences (i.e., trait impulsiveness). Finally, our findings showed that the brain-behavior associations were only evident for trust propensity but not altruistic preferences.

Our findings bolster the assertion that trust propensity is a complex social construct emerging from the interactions among large-scale networks (Bellucci, Chernyak, Goodyear, Eickhoff, & Krueger, 2017; Fehr, 2009). In line with this assumption, a recent study employing resting-state brain-electrical connectivity revealed that trust propensity could be predicted by connections of electrodes located over the frontal and parietal regions (Hahn et al., 2014). However, the EEG-based approach does not allow for a precise location of brain systems contributing to the prediction of trusting propensity. In this regard, our findings provide the evidence to delineate contributions of specific networks in shaping an individual's trust propensity.

Based on a system neuroscience view (Bressler & Menon, 2010), trust emerges from the interactions of psychological systems (i.e., motivation, affect, cognition) that engage key nodes anchored in domain-general large-scale brain networks (Krueger & Meyer-Lindenberg, 2018). The anticipation of reward (motivation, RWN) competed with the risk of betrayal (affect, SAN), which induces uncertainty about the vulnerability of trusting another person. The removal of

uncertainty requires an adoption of a context-based strategy (cognition, CEN) or evaluation of relationship-based trustworthiness (social cognition, DMN); thereby, transforming risk of betrayal to positive expectations of reciprocity.

First, the motivational system of trust involves the RWN that builds on dopaminergic pathways to determine the anticipated reward for trusting another person. As a key node densely interconnected with numerous frontal regions (Yeterian & Van Hoesen, 1978), the caudate is involved in the production of movement; therefore, influencing future planning and decision selection (Joel & Weiner, 1999). Signals from the caudate are predictive of people's decisions to cooperate or to trust (King-Casas et al., 2005; Rilling et al., 2002). The caudate also registers other valuation-related information important for trust decisions, including trustworthiness of others (Dimoka, 2010) and positive feelings of reciprocated trust (Fareri, Chang, & Delgado, 2015; Rilling et al., 2002; Sripada, Angstadt, Liberzon, McCabe, & Phan, 2013).

Second, the affective system of trust engages the SAN to comprise aversive feelings associated with the risk of betrayal by another person. The SAN is consistently identified for the self-related bottom-up saliency detection for regulating social behavior (Bressler &

Menon, 2010). The amygdala, necessary for appropriate social functioning (Adolphs, Tranel, & Damasio, 1998), signals the threat of betrayal based on promoting social vigilance and encoding emotional salience (Engell, Haxby, & Todorov, 2007). Trust increases after damage of the amygdala (Koscik & Tranel, 2011; van Honk, Eisenegger, Terburg, Stein, & Morgan, 2013). The amygdala evaluates incoming social information to enhance (diminish) trust-related behaviors for positive (negative) evaluations—consistent with the literature on the opposite effects of the two hormones oxytocin (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008) and testosterone (Boksem et al., 2013) in balancing trust.

Third, the cognitive system of trust involves the CEN (cognitive control system) to adopt context-based strategies for transforming risk of betrayal to positive expectations of reciprocity. The lateral PFC is a key node of the CEN, which has been consistently associated with top-down cognitive control in adopting goal-directed behavior under changing contexts (Miller & Cohen, 2001). The lateral PFC has been involved in emotion regulation through modulations of bottom-up processes in limbic and subcortical regions (Kober et al., 2010; Lee, Heller, Van Reekum, Nelson, & Davidson, 2012; Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008). These regulatory processes thus might play a key role in transforming risk of betrayal to positive expectations of reciprocity (Fouragnan et al., 2013).

Finally, the cognitive system of trust involves the DMN (social cognition system) to evaluate relationship-based trustworthiness for transforming risk of betrayal to positive expectations of reciprocity. The TP and TPJ are key nodes of the DMN, which have been consistently identified in mentalizing about others to facilitate cooperative decision-making (Frith & Frith, 2006; Lieberman, 2007; Rilling & Sanfey, 2011; Saxe & Kanwisher, 2003). The TP is linked with the representation of abstract social knowledge about other people, such as whether they are trustworthy, that is important for our ability to mentalize (Frith & Frith, 2006; Saxe & Kanwisher, 2003; Zahn et al., 2007). The TPJ is well linked with social cognitive functions, including self-other distinction, perspective-taking, and intentional inferences of others (Van Overwalle, 2009), making it an essential region for inferring and attributing the intentions of others to evaluate relationship-based trustworthiness (Engelmann, Meyer, Ruff, & Fehr, 2018). Trustors with higher perspective-taking tendencies demonstrate greater trust toward others and decrease their trust more drastically after being betrayed (Fett et al., 2014). Sophisticated trustors show higher TPJ activity than naive trustors, consistent with the assumption that they build better mental models about the intentions of their partners (Xiang, Ray, Lohrenz, Dayan, & Montague, 2012). Moreover, TPJ activity increases with age for trusting others—demonstrating a higher sensitivity toward other people's social signals (Fett, Gromann, Giampietro, Shergill, & Krabbendam, 2014). Finally, recent task-based fMRI evidence has shown that the left TPJ is functionally connected with the amygdala and IFG—similar nodes that were also identified in the current study—to mediate human trust behaviors in a context-dependent manner (Engelmann et al., 2018). Notably, both the task-based approach implemented in Engelmann et al. (2018)'s study and current RSFC approach identified the association of left TPJ (but not right TPJ) connectivity profiles with trust behaviors.

Our findings further indicated that trust and associated neural substrates were modulated according to individual differences in

general social preferences (i.e., horizontal collectivism) instead of general risk preferences (i.e., trait impulsiveness). Horizontal collectivism refers to beliefs emphasizing social harmony and interdependence with others (Singelis et al., 1995). These beliefs are closely associated with one's propensity to trust, because trust represents an integral component of positive interpersonal interactions and relationships. Previous evidence has shown that horizontal collectivism is positively associated with self-reported tendencies to trust and cooperate with others (Wagner III, 1995; Wong & Tjosvold, 2006; Zeffane, 2017). Finally, propensity to trust was not predicted by trait impulsiveness—which is closely related to risk preferences in nonsocial contexts (Lauriola, Panno, Levin, & Lejuez, 2014; Romer, 2010; Steinberg, 2010)—supporting previous evidence that trust involves social risk and is distinct from nonsocial risk (Aimone, Ball, & King-Casas, 2015; Baumgartner et al., 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005; Zheng et al., 2017).

Although our findings serve as a starting point, a few limitations exist that future investigations have to address to improve the prediction performance and generalizability of our findings based on RSFC. First, our results indicated that the performance of the predictive models was significantly better than chance level. However, future investigations should improve the accuracy of those models by combining data from different levels of measures (e.g., behavioral and brain measures), modalities (e.g., functional and structural connectivity), and statistical indicators (e.g., mean and variance).

Second, our findings showed that intrinsic functional connectivity profiles only predicted individual differences in trust measured during the first round (i.e., trust propensity) but not trust measured across multiple rounds (i.e., trust dynamics). A similar finding has been found in a recent prediction study—employing resting-state brain-electrical connectivity—where participants played multiple rounds with the same partners providing feedback after each round (Hahn et al., 2014). The authors argue that trust behavior in the first round could have been predicted since trust propensity is considered as a trait reflected in the underlying RSFC, whereas trust behavior averaged across rounds reflects trust dynamic as a state influenced, for example, by social learning (based on the received feedback) not reflected in RSFC. In our study, participants played with different partners with the instruction that feedback will be provided a week later after putative trustees had made their decisions, considering the first round of their trust behavior would reflect their propensity to trust as measured in the identified RSFC. In contrast, considering the average trust behavior across rounds could reflect trust dynamics, where participants, for example, strategizing how to play the game differently to maximize their outcome across all games. Future studies are needed to disentangle those different trust dynamics (e.g., social learning, strategizing) and their underlying neural underpinnings based on task-based and task-free functional connectivity.

Finally, we investigated whether RSFC predicts individual variations of trust propensity at a single time point, but future studies should further investigate whether the identified RSFC pattern is a temporally stable index of trust propensity over a longer time interval—representing the underlying neural patterns of an individual's phenotype (Peysakhovich, Nowak, & Rand, 2014).

Despite these limitations, our study provided the first evidence that the functional interplay among distributed networks predicts individual variation in the propensity of trust. The contributing nodes and edges of the predictive networks—representing the motivational, affective, and cognitive components of trust—do not work separately, but extensively interact with each other. In this regard, our data-driven approach provides a more holistic measure of trust and offers a novel tool to characterize the neural underpinnings, highlighting its potential use as an objective neuromarker (Kristensen & Sandberg, 2017; Kropotov, 2016) for future trust investigations in mental health disorders that are characterized by mistrust in close social relationships (Fett et al., 2012; Sripada et al., 2009; Unoka et al., 2009).

ACKNOWLEDGMENTS

This work was supported by the National Postdoctoral Program for Innovative Talents (BX201600019), the China Postdoctoral Science Foundation (2017M610055), and the National Natural Science Foundation of China (31530031). The authors thank Dr. James Rilling for his helpful comments on an early version of the manuscript.

CONFLICT OF INTERESTS

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

ORCID

Chunliang Feng <https://orcid.org/0000-0001-5595-2146>

REFERENCES

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, 393, 470–474.
- Aimone, J., Ball, S., & King-Casas, B. (2015). The betrayal aversion elicitation task: An individual level betrayal aversion measure. *PLoS One*, 10, e0137491.
- Alarcon, G. M., Lyons, J. B., Christensen, J. C., Bowers, M. A., Klosterman, S. L., & Capiola, A. (2018). The role of propensity to trust and the five factor model across the trust process. *Journal of Research in Personality*, 75, 69–82.
- Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, 58, 639–650.
- Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., ... Kane, M. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences*, 115, 1087–1092.
- Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S. B., & Krueger, F. (2017). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. *Human Brain Mapping*, 38, 1233–1248.
- Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, 28, 168–175.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122–142.
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34, 537–541.
- Boksem, M. A., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., ... Sanfey, A. G. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychological Science*, 24, 2306–2314.
- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, 14, 277–290.
- Camerer, C., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica: Journal of the Econometric Society*, 56, 1–36.
- Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences*, 105, 3721–3726.
- Chu, C., Ni, Y., Tan, G., Saunders, C. J., & Ashburner, J. (2011). Kernel regression for fMRI pattern prediction. *NeuroImage*, 56, 662–673.
- Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, 178, 622–637.
- Cui, Z., Su, M., Li, L., Shu, H., & Gong, G. (2018). Individualized prediction of reading comprehension ability using gray matter volume. *Cerebral Cortex*, 28, 1656–1672.
- Cui, Z., Xia, Z., Su, M., Shu, H., & Gong, G. (2016). Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. *Human Brain Mapping*, 37, 1443–1458.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8, 1611–1618.
- Dimoka, A. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly*, 34, 373–396.
- Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, 20, 425–443.
- Ecker, C., Marquand, A., Mourao-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., ... Murphy, D. G. (2010). Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 30, 10612–10623.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, 19, 1508–1519.
- Engelmann, J. B., Meyer, F., Ruff, C. C., Fehr, E. (2018). The neural circuitry of emotion-induced distortions of trust. *bioRxiv*:129130.
- Erus, G., Battapady, H., Satterthwaite, T. D., Hakonarson, H., Gur, R. E., Davatzikos, C., & Gur, R. C. (2015). Imaging patterns of brain development and their relationship to cognition. *Cerebral Cortex*, 25, 1676–1684.
- Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *The Journal of Neuroscience*, 35, 8170–8180.
- Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, 7, 235–266.
- Feng, C., Azarian, B., Ma, Y., Feng, X., Wang, L., Luo, Y. J., & Krueger, F. (2017). Mortality salience reduces the discrimination between in-group and out-group interactions: A functional MRI investigation using multi-voxel pattern analysis. *Human Brain Mapping*, 38, 1281–1298.
- Feng, C., Cui, Z., Cheng, D., Xu, R., & Gu, R. (2018). Individualized prediction of dispositional worry using white matter connectivity. *Psychological Medicine*. <https://doi.org/10.1017/S0033291718002763>. [Epub ahead of print]
- Feng, C., Yuan, J., Geng, H., Gu, R., Zhou, H., Wu, X., & Luo, Y. (2018). Individualized prediction of trait narcissism from whole-brain resting-state functional connectivity. *Human Brain Mapping*, 39, 3701–3712.
- Feng, C., Zhu, Z., Gu, R., Wu, X., Luo, Y.-J., & Krueger, F. (2018). Resting-state functional connectivity underlying costly punishment: A machine learning approach. *Neuroscience*, 385, 25–37.
- Fett, A. K., Gromann, P. M., Giampietro, V., Shergill, S. S., & Krabbendam, L. (2014). Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Social Cognitive and Affective Neuroscience*, 9, 395–402.
- Fett, A. K., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S. J., Yakub, F., & Krabbendam, L. (2014). Trust and social reciprocity in adolescence—a matter of perspective-taking. *Journal of Adolescence*, 37, 175–184.

- Fett, A. K., Shergill, S. S., Joyce, D. W., Riedl, A., Strobel, M., Gromann, P. M., & Krabbendam, L. (2012). To trust or not to trust: The dynamics of social interaction in psychosis. *Brain*, 135, 976–984.
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18, 1664–1671.
- Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., & Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 33, 3602–3611.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 9673–9678.
- Franke, K., Ziegler, G., Klöppel, S., Gaser, C., & Initiative, A. S. D. N. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, 50, 883–892.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35, 346–355.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50, 531–534.
- Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85, 11–26.
- Gianotti, L. R., Lobmaier, J. S., Calluso, C., Dahinden, F. M., & Knoch, D. (2017). Theta resting EEG in TPJ/pSTS is associated with individual differences in the feeling of being looked at. *Social Cognitive and Affective Neuroscience*, 13, 216–223.
- Gianotti, L. R., Nash, K., Baumgartner, T., Dahinden, F. M., & Knoch, D. (2018). Neural signatures of different behavioral types in fairness norm compliance. *Scientific Reports*, 8, 10513.
- Gong, Q., Li, L., Du, M., Pettersson-Yeo, W., Crossley, N., Yang, X., ... Mechelli, A. (2014). Quantitative prediction of individual psychopathology in trauma survivors using resting-state FMRI. *Neuropsychopharmacology*, 39, 681–687.
- Hahn, T., Notebaert, K., Anderl, C., Teckentrup, V., Kaßbecker, A., & Windmann, S. (2014). How to trust a perfect stranger: Predicting initial trust behavior from resting-state brain-electrical connectivity. *Social Cognitive and Affective Neuroscience*, 10, 809–813.
- Hsu, W., Rosenberg, M., Scheinost, D., Constable, R., & Chun, M. (2018). Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Social Cognitive and Affective Neuroscience*, 13, 224–232.
- Joel, D., Weiner, I. (1999). Striatal contention scheduling and the split circuit scheme of basal ganglia-thalamocortical circuitry: from anatomy to behaviour. In: *Brain dynamics and the striatal complex*. Amsterdam, Harwood Academic Publishers.
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32, 865–889.
- Jung, W. H., Lee, S., Lerman, C., & Kable, J. W. (2018). Amygdala functional and structural connectivity predicts individual risk tolerance. *Neuron*, 98, 394–404 e4.
- Kable, J. W., & Levy, I. (2015). Neural markers of individual differences in decision-making. *Current Opinion in Behavioral Sciences*, 5, 100–107.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59, S285–S300.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78–83.
- Kober, H., Mende-Siedlecki, P., Kross, E. F., Weber, J., Mischel, W., Hart, C. L., & Ochsner, K. N. (2010). Prefrontal–striatal pathway underlies cognitive regulation of craving. *Proceedings of the National Academy of Sciences*, 107, 14811–14816.
- Koscik, T. R., & Tranel, D. (2011). The human amygdala is necessary for developing and expressing normal interpersonal trust. *Neuropsychologia*, 49, 602–611.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435, 673–676.
- Kristensen, S.B., Sandberg, K. (2017). Is whole-brain functional connectivity a neuromarker of sustained attention? Comment on Rosenberg et al. (2016). *bioRxiv*:216697.
- Kropotov, J. D. (2016). *Functional neuromarkers for psychiatry: Applications for diagnosis and treatment*. San Diego, CA, USA: Academic Press.
- Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., ... Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104, 20084–20089.
- Krueger, F., & Meyer-Lindenberg, A. (2018). Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends in Neurosciences*. S0166-2236(18)30275-3. <https://doi.org/10.1016/j.tins.2018.10.004>. [Epub ahead of print]
- Krueger, F., Parasuraman, R., Iyengar, V., Thornburg, M., Weel, J., Lin, M., ... Lipsky, R. (2012). Oxytocin receptor genetic variation promotes human trust behavior. *Frontiers in Human Neuroscience*, 6, 4.
- Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, 27, 20–36.
- Lee, H., Heller, A. S., Van Reekum, C. M., Nelson, B., & Davidson, R. J. (2012). Amygdala–prefrontal coupling underlies individual differences in emotion regulation. *NeuroImage*, 62, 1575–1581.
- Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, 58, 259–289.
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, 98, 11832–11835.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Nash, K., Gianotti, L. R., & Knoch, D. (2015). A neural trait approach to exploring individual differences in social preferences. *Frontiers in Behavioral Neuroscience*, 8, 458.
- Nash, K., Knoch, D. (2016). Individual differences in decision-making: A neural trait approach to study sources of behavioral heterogeneity. In: *Neuroeconomics* (pp. 191–209). Springer.
- Nishina, K., Takagishi, H., Inoue-Murayama, M., Takahashi, H., & Yamagishi, T. (2015). Polymorphism of the oxytocin receptor gene modulates behavioral and attitudinal trust among men but not women. *PLoS One*, 10, e0137089.
- Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, 51, 768–774.
- Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*, 5, 4939.
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59, 2142–2154.
- Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, 35, 395–405.
- Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, 62, 23–48.
- Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: Implications for prevention. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 52, 263–276.
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19, 165–171.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, 19, 1835–1842.
- Seppänen, R., Blomqvist, K., & Sundqvist, S. (2007). Measuring inter-organizational trust—A critical review of the empirical research in 1990–2003. *Industrial Marketing Management*, 36, 249–265.
- Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based

- predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12, 506–518.
- Shen, X., Papademetris, X., & Constable, R. T. (2010). Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. *NeuroImage*, 50, 1027–1035.
- Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Group-wise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82, 403–415.
- Shin, H. H., & Park, T. H. (2005). Individualism, collectivism and trust. *International Review of Public Administration*, 9, 145–161.
- Singelis, T. M., Triandis, H. C., Bhawuk, D. P., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29, 240–275.
- Snyder, A. Z., & Raichle, M. E. (2012). A brief history of the resting state: The Washington university perspective. *NeuroImage*, 62, 902–910.
- Sripada, C., Angstadt, M., Liberzon, I., McCabe, K., & Phan, K. L. (2013). Aberrant reward center response to partner reputation during a social exchange game in generalized social phobia. *Depression and Anxiety*, 30, 353–361.
- Sripada, C. S., Angstadt, M., Banks, S., Nathan, P. J., Liberzon, I., & Phan, K. L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport*, 20, 984–989.
- Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 52, 216–224.
- Steinberg, L., Sharp, C., Stanford, M. S., & Tharp, A. T. (2013). New tricks for an old measure: The development of the Barratt impulsiveness scale-brief (BIS-brief). *Psychological Assessment*, 25, 216–226.
- Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1, 211–244.
- Unoka, Z., Seres, I., Áspán, N., Bódi, N., & Kéri, S. (2009). Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*, 23, 399–409.
- van Honk, J., Eisenegger, C., Terburg, D., Stein, D. J., & Morgan, B. (2013). Generous economic investments after basolateral amygdala damage. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 2506–2510.
- Van Lange, P. A., Rockenbach, B., & Yamagishi, T. (2017). *Trust in social dilemmas*. New York, NY, US: Oxford University Press.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30, 829–858.
- Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145, 166–179.
- Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59, 1037–1050.
- Wagner, J. A., III. (1995). Studies of individualism-collectivism: Effects on cooperation in groups. *Academy of Management Journal*, 38, 152–173.
- Wang, L., Lu, X., Gu, R., Zhu, R., Xu, R., Broster, L. S., & Feng, C. (2017). Neural substrates of context-and person-dependent altruistic punishment. *Human Brain Mapping*, 38, 5535–5550.
- Wang, Y., Fan, Y., Bhatt, P., & Davatzikos, C. (2010). High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. *NeuroImage*, 50, 1519–1535.
- Wong, A., & Tjosvold, D. (2006). Collectivist values for learning in organizational relationships in China: The role of trust and vertical coordination. *Asia Pacific Journal of Management*, 23, 299–317.
- Xiang, T., Ray, D., Lohrenz, T., Dayan, P., & Montague, P. R. (2012). Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Computational Biology*, 8, e1002841.
- Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., ... Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76, 183–201.
- Yan, C.-G., Wang, X.-D., Zuo, X.-N., & Zang, Y.-F. (2016). DPABI: Data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, 14, 339–351.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12, 1100–1122.
- Yeterian, E. H., & Van Hoesen, G. W. (1978). Cortico-striate projections in the rhesus monkey: The organization of certain cortico-caudate connections. *Brain Research*, 139, 43–63.
- Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6430–6435.
- Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One*, 2, e1128.
- Zeffane, R. (2017). Gender, individualism–collectivism and individuals' propensity to trust: A comparative exploratory study. *Journal of Management & Organization*, <https://doi.org/10.1017/jmo.2017.57>.
- Zheng, H., Wang, S., Guo, W., Chen, S., Luo, J., Ye, H., & Huang, D. (2017). Enhancing the activity of the DLPFC with tDCS alters risk preference without changing interpersonal trust. *Frontiers in Neuroscience*, 11, 52.
- Zhong, S., Monakhov, M., Mok, H. P., Tong, T., San Lai, P., Chew, S. H., & Ebstein, R. P. (2012). U-shaped relation between plasma oxytocin levels and behavior in the trust game. *PLoS One*, 7, e51095.
- Zuo, X.-N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., ... Milham, M. P. (2010). The oscillating brain: Complex and reliable. *NeuroImage*, 49, 1432–1445.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lu X, Li T, Xia Z, et al. Connectome-based model predicts individual differences in propensity to trust. *Hum Brain Mapp*. 2019;40:1942–1954. <https://doi.org/10.1002/hbm.24503>