



Hierarchical Neural Prediction of Interpersonal Trust

Yiwen Wang^{1,2} · Xue Yang^{1,2} · Zhenpeng Tang¹ · Shaobei Xiao³ · Johannes Hewig⁴

Received: 2 March 2020 / Accepted: 29 August 2020 / Published online: 9 February 2021
© Center for Excellence in Brain Science and Intelligence Technology, CAS 2021

Abstract Exploring neural markers that predict trust behavior may help us to identify the cognitive process underlying trust decisions and to develop a new approach to promote interpersonal trust. It remains unknown how trust behavior may be predicted early in the decision process. We used electrophysiology to sample the brain activity while participants played the role of trustor in an iterative trust game. The results showed that during the trust generation stage, the trust condition led to higher frontocentral beta band activity related to cognitive inhibition compared to the distrust condition (item level). Moreover, individuals with higher frontocentral beta band activity were more likely to perform trust choices at the single-trial level (individual level). Furthermore, after receiving reciprocity feedback on trial_{n-1}, compared to the betrayal feedback and the distrust choice, the frontocentral beta band oscillation had a stronger predictive

effect regarding trust choices on trial_n. These findings indicate that beta band oscillations during the decision generation stage contribute to subsequent trust choices.

Keywords Interpersonal trust · hierarchical prediction · trust game · beta band oscillation

Introduction

Trust is a critical social component of human interactions that help individuals cooperate with others [1, 2]. Predicting trust, as one of the most critical issues in the field of trust behavior, can assist us in the future to promote engagement better in social activities and economic transactions. To date, a growing branch of research examining brain neuroimaging, peptide hormones, and personality trait measurement strives to predict social trust in individuals [3–5]. Despite this rich literature on prediction, there is a lack of evidence about the use of oscillatory brain activity to predict individuals' subsequent social trust behavior during social interaction. This is particularly relevant since real-time and task-based brain oscillations have high temporal precision and may have relatively high accuracy and applicability when predicting trust choices [6, 7]. Thus, it seems necessary to predict trust using a neural index before trust behavior is performed, and such evidence may provide a new approach to promoting interpersonal trust. The approach of using electroencephalography (EEG) signals to predict trust is flexible and possible in real time (e.g., [8–10]), and it also provides a basis for understanding the cognitive processes underlying interpersonal trust.

To predict trust choices, it is critical to clarify the cognitive process of trust decisions from the perspective of

✉ Yiwen Wang
wangeven@126.com

✉ Zhenpeng Tang
zhenpt@126.com

✉ Shaobei Xiao
xiaoshb@hainnu.edu.cn

✉ Johannes Hewig
hewig@psychologie.uni-wuerzburg.de

¹ School of Economics and Management, Fuzhou University, Fuzhou 350108, China

² Institute of Psychological and Cognitive Sciences, Fuzhou University, Fuzhou 350108, China

³ School of Psychology, Hainan Normal University, Haikou 571158, China

⁴ Department of Psychology, Julius Maximilians University Würzburg, 97070 Würzburg, Germany

information processing, which consists of three interrelated stages in typical trust game paradigms [11–13]. Among the three stages of information processing, the first is decision generation in which the trust scenario is presented. During this stage, individuals spontaneously generate a tendency to favor one choice alternative over others. The second step is the decision output where the trust decision is executed. During this process, a single behavioral response is produced, and neural correlates typically reflect the categorical output of the decision process. The third step is the outcome evaluation that occurs after the appearance of outcome feedback. During the outcome evaluation stage, the brain encodes the success or failure of the behavior, and individuals rely on this information to guide future decisions.

Recently, an increasing number of studies have suggested that the cognitive process during the decision generation stage significantly predicts subsequent decision outputs of an individual [14, 15]. Specifically, studies have attempted to explore the cognitive processing at the stage of decision generation and decision output. These findings have indicated that when an individual is thinking about the current social context without a direct behavioral decision-making output (e.g., stage of decision generation), the elicited neural pattern resembles that at the stage of behavioral decision output. More importantly, the cognitive process during the decision generation stage is similar to subsequent social behavioral outputs [15]. Therefore, the described proximity between the stage of decision generation and decision outcome makes it highly plausible that trust can be predicted during the trust decision generation stage (i.e., before the trust decision output). Given the above, we hypothesized that there would be a neural oscillation marker during the stage of decision generation to predict the subsequent trust choices of individuals.

Previous studies have mainly focused on the decision output stage and suggested that cognitive inhibition mechanisms play an important role in distinguishing trust and distrust choices [16–18]. From the perspective of cognitive inhibition, the neural process of cognitive inhibition contributes to maintaining choices aligned with reliable prior beliefs when beliefs momentarily conflict with observations. Considering the reciprocity rate from the opponent in the trust game [11, 16], the relative income of distrust choices is 10 points, and the relative income of trust choices is equal to 10 points [i.e., 20 (opponent reciprocity) \times 50% (probability) + 0 (opponent betrayal) \times 50% (probability)]. There is no long-term cost or benefit of trusting from the perspective of betrayal aversion, so individuals would hold the *status quo* as the inherent option tendency to protect their own interest [19–22]. As above, it is reasonable to suppose that during the trust generation stage, individuals maintain “opponent untrustworthiness”

prior beliefs under such circumstances, and the trust choices would be accompanied more by beta band oscillations related to cognitive inhibition during the decision generation stage than the distrust choices.

An increasing amount of evidence has suggested that stronger cognitive inhibition is associated with increased frontocentral beta band activity [23, 24]. In trust tasks, previous studies have further demonstrated that the cognitive inhibition-related beta band oscillations are associated with relatively more trust behavior [16, 25]. For example, Fu *et al.* using the iterative trust game suggested that the higher the frontocentral beta band power in the time interval of ~ 200 – 400 ms, the higher the probability of a trust choice [16]. This study demonstrated that when an individual thinks the opponent is untrustworthy, cognitive inhibition resources are used to perform the trust behavior. Also, the study by Wang *et al.*, in which “trust” was the advantageous tendency, the distrust choices that conflicted with prior beliefs showed greater frontocentral beta band power related to cognitive inhibition in the time interval about 250–350 ms compared to trust choices [18]. We inferred that the cognitive inhibition-related beta band power during the trust generation stage, in which individuals start to evaluate the trust situation, would be a potential neural indicator for predicting individual trust decisions.

Furthermore, although previous studies have repeatedly demonstrated that outcome feedback influences subsequent trust choices in the iterative trust game (i.e., tit-for-tat strategy: individuals prone to trust more after receiving positive reciprocity feedback, and to distrust more after receiving negative betrayal feedback [26–28]), it remains unclear how the outcome feedback on trial_{n-1} moderates the predictive effect of cognitive inhibition on the subsequent decision in trial_n. Consistent with the model of decision circuits [13], previous studies have suggested that the trust choices (at the behavioral and neural levels) are influenced by the outcome feedback on the prior trial [27–29]. At the behavioral level, studies (e.g., [27]) have found that participants adjust their trust decisions over rounds, and the received outcome feedback is a strong predictor of subsequent trust choices. At the neural level, Fouragnan and colleagues found that when individuals interact with a cooperative counter-partner, the prefrontal cortex activity related to cognitive inhibition positively predicts the rates of continuing choices to trust after receiving betrayal outcome feedback rather than cooperative outcome feedback [29]. The results of the outcome feedback mediate the correlation between the prefrontal cortex and subsequent trust choices, which might suggest that the cognitive inhibition-related prefrontal cortex activity contributes to maintaining inherent prior beliefs (i.e., trusting the cooperative counter-partner) when beliefs momentarily conflict

with the betrayal feedback observation. As noted above, the outcome feedback on trial_{n-1} might moderate the degree of cognitive inhibition-related oscillatory beta band activity when predicting subsequent trust choices on trial_n.

In summary, the present study had two major purposes by analyzing event-related spectral perturbations (ERSPs) during presentation of the binary trust tree (the stage of decision generation) in the trust game. One goal was to investigate whether the oscillatory beta band power related to cognitive inhibition predicted subsequent trust choices at both the item and the individual level. At the item level, we predicted that during the decision generation stage, there would be greater cognitive inhibition-related beta band power among the trust trials than the distrust trials, corroborating this activity as a potential neural indicator of individual trust choices. Moreover, at the individual level, we predicted beta band oscillations could positively predict trust choices, reflecting a specific neural activity to predict individual trust behavior. The other goal was to investigate whether the outcome feedback modulated the prediction effect of the cognitive inhibition-related beta band oscillation on individual trust choices. We hypothesized that, at the trial-by-trial level, the frontocentral beta band power would be more strongly predictive of trust choices after receiving trust reciprocity feedback than after receiving betrayal feedback, compared to making a distrust choice.

Materials and Methods

Statistical Power and Participants

Because we were uncertain as to the effect sizes we may find in the current study, we referred to the sample size of the most analogous study in the literature, which was the demonstration of the main effect of trust decision choices by Wang *et al.* [17]. We used their effect size ($\eta_p^2 = 0.18$) to estimate our sample size using G*Power (version 3.1; [30]). This analysis suggested we collected at least 17 participants to obtain 80% power. Conservatively, 20 undergraduate students were recruited from the local university. All participants were right-handed and reported normal or corrected-to-normal vision; they were screened for neurological conditions and had no history of medical, neurological, or psychiatric diagnoses; and they provided informed consent and participated in financial compensation in the EEG experiment. All procedures were in accordance with the 1964 Declaration of Helsinki and were approved by the local Ethics Committee. During the data analysis process, two participants were excluded because of excessive artifacts in the electrophysiological recordings. As a result, 18 participants (11 females, mean age \pm

SD, 20.33 ± 2.00 years; range, 18 to 25 years) were included in the analysis.

Behavioral Task

The behavioral task adopted the multi-shot version trust game wherein a trustor (participant) and a trustee (unacquainted with each other) complete the task together [11]. In each round, both trustor and trustee endow 10 game points as the initial funding. First, the trustor is required to decide whether to invest all 10 points to the trustee or keep the endowment. If the trustor chooses to keep, the current round ends and both players receive 10 points. If the trustor chooses to invest, the endowment is tripled to 30 points and then transferred to the trustee. That is, the trustee owns 40 points (tripled points plus his/her own 10 points). Next, the trustee chooses whether to keep the 40 points or return half of the points to the trustor. For the former, the trustee endows all 40 points and the trustor loses his initial endowments; and for the latter, both players would endow 20 game points. Considering the uncertainty of the trustee's reciprocity behavior, the trustor's investment choice reflects his/her willingness to be vulnerable to the trustee's decision, which corresponds to the operational definition of trust [31].

Stimuli and Procedures

Participants entered the laboratory, and were given written instructions explaining the experiment rules as well as the payment procedures among this experiment. All participants were informed that their payoff would rely on the cumulative points earned during the experiment. To ensure the experiment instructions were understood, the participants restated the rules briefly on request. Before the formal experiment, participants performed a practice task lasting about 4 min for which they did not receive payment.

Then, participants were seated in a quiet dimly-lit room about 80 cm from a computer screen. Participants completed 100 rounds of the trust game and their brain activity was recorded *via* electroencephalography (EEG). Participants were not told the exact rounds of the trust game before completing the experimental task. In each round, a "binary trust tree" was presented for 1500 ms on the central screen (Fig. 1). Then a fixation cross ("+") was presented for a random duration (800–1000 ms) on a blank screen. After the fixation cross, the decision options showing "10" and "30" were presented, and participants were required to choose one of the options by pressing "1" or "3" on the keyboard using the right index or middle finger. To control the response patterns, 10 participants were presented in a layout of the "binary game tree" as illustrated in Fig. 1, and the others in its mirrored layout. If

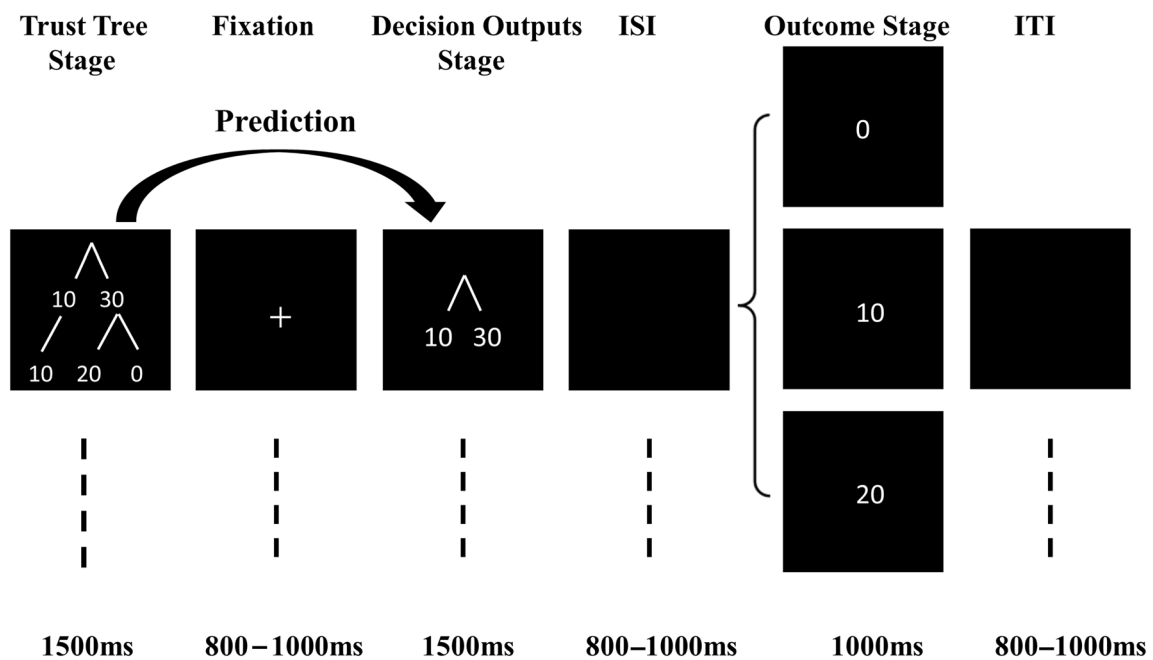


Fig. 1 Illustration of one trial in the iterative trust game. Presentation of the “binary trust tree” (decision generation stage) represents the beginning of the current round. At the decision output stage, the invest endowment decision (cued by “30” on the slide) represents trust, or else, the *status quo* decision (cued by “10” on the slide)

represents distrust. At the outcome evaluation stage, 0 points (cued by “0”) represents a trust betrayal outcome, 10 points (cued by “10”) represents the trustors’ distrust choice, and 20 points (cued by “20”) represents a trust reciprocity outcome.

a participant chose to invest, a 1500 ms blank screen was presented while the trustee was asked to make a reciprocity decision. Finally, the trustor’s outcome (reciprocity: 0 points or betrayal: 20 points) of the current round was displayed for 1000 ms. If the participant chose to keep, the outcome (10 points) was displayed directly for 1000 ms and the current round finished. Once the participant failed to make a decision during the 1500 ms, the current round was restarted in order to ensure the overall number of points was constant.

Note that the trustees were lab assistants (1 female and 1 male, 24 and 23 years old respectively), and the participants (trustors) interacted with the trustee of the same gender (experimental assistant). First, the trustee was set to the same person to avoid trust decisions affected by the characteristics of attractiveness and trustworthiness (e.g., [32]). Second, the identical gender was chosen to avoid the influence of gender stereotypes on trust decisions (e.g., [33]). In fact, the trustees’ reciprocal behaviors were based on a pre-programmed procedure that the reciprocity rates were 50%. The above setting ensured that all participants were in an identical social situation.

Participants received a fixed compensation (25 CNY) for joining this experiment. In addition, their compensation was linked to their cumulative earnings during the experiment [17, 18]. Overall, the participants received 50 CNY on average at the end of the experiment.

EEG recording and Analysis

Raw EEG data were recorded using an elastic cap with 64 electrodes (Neuroscan, Neurosoft Labs, Inc. Sterling, USA). All electrodes were placed according to the international 10-20 system. Vertical electrooculogram (EOG) activity was recorded with electrodes placed supra- and infra-orbitally at the left eye. Horizontal EOG activity was recorded with electrodes placed on the left and right orbital rim. All inter-electrode impedances were maintained below 5 k Ω . The EEG and EOG were amplified with a DC~100 Hz bandpass and continuously sampled at 1000 Hz/channel for offline analysis.

Off-line EEG data were preprocessed and analyzed using the EEGLAB toolbox ([34], version 13.4.4b) in the MatLab (version R2014a) environment. EEG data were referenced to the averaged bilateral mastoids and then filtered using Butterworth filters with half-power cutoffs at 0.01 and 35 Hz. Independent component analysis (Runica algorithm) was subsequently performed to identify and remove components associated with eye movements and eye-blinks. All trials with amplifier clipping artifacts or EEG voltages exceeding the threshold of $\pm 80 \mu\text{V}$ were marked as “noise” and excluded from subsequent analysis. Channels with >25% of the overall trials marked “noise” were regarded as bad channels. Bad channels were

interpolated using the spherical option of the EEGLAB function for channel interpolation [35].

Because of the purpose of the study, we analyzed the EEG epochs that were time-locked to the presentation of the “binary game tree” (i.e., trust decision generation stage). Single-trial epochs comprised a 2300 ms window, including 800 ms prior to the “binary game tree” onset. On average, there were 91.50 ± 4.81 effective trials per participant: 40.55 ± 12.62 trust trials (range, 23 to 68) per participant, and 50.94 ± 13.14 distrust trials (range, 23 to 70) per participant. The EEG data were down-sampled to 500 Hz. Oscillatory power, calculated as a function of time and frequency, was obtained from the single-trial EEG epochs per channel using the Morlet-based wavelet transform procedure in the EEGLAB toolbox. The single-trial time-frequency representations were used among 3 Hz and 35 Hz between -800 ms and 1500 ms, and then were averaged to obtain averaged time-frequency representations, which were adopted to identify modulations of the ongoing EEG rhythms (ERSPs) and yield the corresponding time-frequency maps [36]. To avoid edge effects, power values were normalized with respect to the -400 ms to -100 ms pre-stimulus baseline interval and converted to decibels [$10 \times \log(\mu V^2)$].

To identify brain oscillation responses that reliably predict trust choices, we took the following steps of analysis. First, EEGs in the trust decision generation stage were divided into two conditions (trust and distrust conditions) according to the subsequent observed trust choices in the decision output stage (at the item level). In this step, we first performed a point-by-point statistical analysis (permutation test) and then confirmed the results using region-of-interest (ROI)-based statistical analyses in the time-frequency domain. Second, we used the logit model to test whether the defined beta band oscillations significantly predicted trust choices (at the individual level). Furthermore, we carried out a cross-validation analysis to verify the robust generalizability of the prediction of trust choices by oscillatory brain activity. Finally, to address the modulatory effects of the outcome feedback, we used the defined neural marker to explore how the outcome feedback modulated the observed prediction effects through a trial-by-trial analysis (at the trial-by-trial level).

In specifically, to determine whether the ERSP distinguishes between trust and distrust choices at the item level we used the following analyses. The oscillatory power was computed as the mean spectral power by averaging the frequency spectra of all participants with the trusting choice and distrusting choice. Then, permutation tests were implemented on the ERSP data of the trust and distrust conditions using the “statcond” function of EEGLAB in order to test condition differences. The Bonferroni

correction was conducted for multiple comparisons [18]. Through the permutation tests, we defined the time-frequency ROI (TF-ROI) and spatial ROI (S-ROI) that significantly distinguished trust and distrust. Next, we adopted the acquired specific single-trial oscillatory power (including TF-ROI and S-ROI) and computed the logit model at the individual level to test whether the defined oscillatory power predicted the individual trust choices using single-trial Hierarchical Linear Model (HLM) analysis (at the individual level). Then, we used k-fold (10-fold) cross-validation analysis to establish the robust generalizability of brain activation to predict trusting [37, 38]. Furthermore, to verify that the prediction effect was moderated by the outcome feedback on trial_{n-1} , based on the data at the individual level, we added the data of outcome feedback on trial_{n-1} , and adopted the two-level logit model to test the moderation effects of outcome feedback on trial_{n-1} .

Results

Behavioral Results

1. Trial analysis of behavioral trust choices

On average, the overall percentages of trust (invest) and distrust (keep) were $45.50 \pm 12.78\%$ and $54.50 \pm 12.78\%$, respectively. There was no significant difference between the trust rates and the distrust rates ($t = -1.494$, $P = 0.154$, Cohen $d' = 0.352$). The overall reaction times (RTs) of trust and distrust among the participants were 519 ± 120 ms and 520 ± 151 ms, respectively. There was no significant difference in RTs between trusting and distrusting choices ($t = 0.070$, $P = 0.945$, Cohen $d' = 0.016$).

Furthermore, we performed a trial-by-trial analysis to explore the effect of prior outcome feedback (trust reciprocity feedback/trust betrayal feedback/distrust choice) on trusting/distrusting choices. Note that in this trial-by-trial analysis, the first trial of each participant was excluded due to the lack of prior outcome feedback information. One-way ANOVA revealed a significant effect of trial_{n-1} outcome feedback on trial_n trust choices ($F(2, 51) = 7.067$, $P = 0.002$, $\eta_p^2 = 0.217$). The probability of trust choices was significantly higher after receiving reciprocity feedback than betrayal feedback ($M_{\text{reciprocity}} = 0.622 \pm 0.272$; $M_{\text{betrayal}} = 0.351 \pm 0.195$; $P = 0.002$, Cohen $d' = 1.181$) and distrust choice ($M_{\text{distrust}} = 0.438 \pm 0.189$; $P = 0.047$, Cohen $d' = 0.813$). However, there was no significant difference in trust choices between receiving betrayal feedback and preceding distrust choices ($P = 0.729$, Cohen $d' = 0.470$) (Fig. 2A).

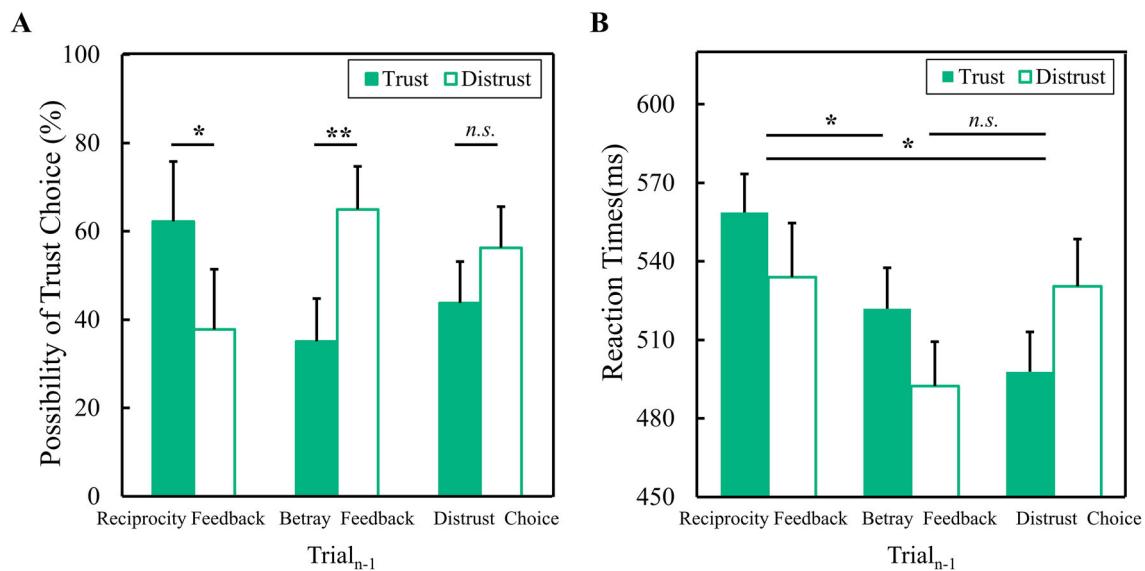


Fig. 2 Behavioral results on trial_n after receiving different outcome feedback on trial_{n-1}. **A** Probability of trust choice. **B** Average RTs of trust and distrust choice. * $P < 0.05$, n.s., no significant difference, $P > 0.05$.

Concerning the effects of the trial_{n-1} feedback on trial_n trust choices from the aspect of individual RTs, we ran a 3 (trial_{n-1}: trust reciprocity feedback *vs* trust betrayal feedback *vs* distrust choice) \times 2 (trial_n: trust choice *vs* distrust choice) within-subject repeated-measures ANOVA and found significant main effects of prior outcome feedback ($F(2,34) = 5.277$, $P = 0.010$, $\eta_p^2 = 0.237$). The *post hoc* test showed that after receiving the reciprocity feedback participants took more time to make a decision compared to after receiving betrayal feedback ($M_{\text{reciprocity}} = 546 \pm 139$ ms; $M_{\text{betrayal}} = 507 \pm 132$ ms; $P = 0.013$, Cohen $d' = 0.779$) and under the distrust choice condition ($M_{\text{distrust}} = 514 \pm 134$ ms; $P = 0.017$, Cohen $d' = 0.738$). In addition, the two latter conditions had no significant difference in RTs ($P = 1.000$, Cohen $d' = 0.096$). Furthermore, there was no significant interaction effect of the trial_{n-1} feedback and the trial_n trust choice ($F(2, 34) = 3.073$, $P = 0.059$, $\eta_p^2 = 0.153$), and no main effect of the trial_n trust choice ($F(1, 17) = 0.162$, $P = 0.693$, $\eta_p^2 = 0.009$) (Fig. 2B).

2. Time series analysis of behavioral trust choices

To further explore the characteristics of the time series of trust choices, a longitudinal analysis was used to account for the trial-related effects through the current trust game. Fig. 3 shows the dynamic trust rate of individuals; the fitted curve was $y = -0.064\ln(x) + 0.681$.

Predictive Analysis of ERSP in the Decision Generation Stage

1. Beta band power predicts trust choices at the item level
We used the participants' trusting and distrusting choices during the decision output stage to distinguish the corresponding neural oscillation differences in the decision generation stage. This analysis at the item level examined whether there was a specific time-frequency of brain oscillation in the decision generation stage that can predict subsequent trust choices. Based on the previous studies [16–18] and our experimental hypothesis, we focused on the brain regions of the frontal electrodes, which are related to the processing of cognitive inhibition. Permutation tests were implemented in the “statcond” function and at a significance level of 0.05. Visual inspection of the pattern of neural oscillation differences between the trusting and distrusting conditions revealed that the beta band oscillations (14–21 Hz) from 200 to 400 ms were significantly higher in the trusting condition than in the distrusting condition at the F1, F3, FC1, and FC3 electrodes, as expected based on previous research (Fig. 4). The time-frequency representations in the S-ROIs (F1, F3, FC1, and FC3) are illustrated in Fig. 4A. The TF-ROIs in the beta band (8–12 Hz, 200–400 ms) were defined with the most pronounced main effects (Fig. 4A). The different scalp topographies of ERSP magnitudes for trust and distrust trials in the TF-ROIs are illustrated in Fig. 4B. One-way ANOVA showed a significant main effect between the trust trials and the distrust trials ($F(1, 17) = 17.386$,

Fig. 3 Individual trust rates over the course of the trust game. The graph shows the fraction of trusting subjects for each round in the iterative trust game (dots, average trust rate in a particular round (ranging from 1 to 100); shaded area, 95% confidence interval; colored line, time trend; three shades of gray in the background, the three time phases).

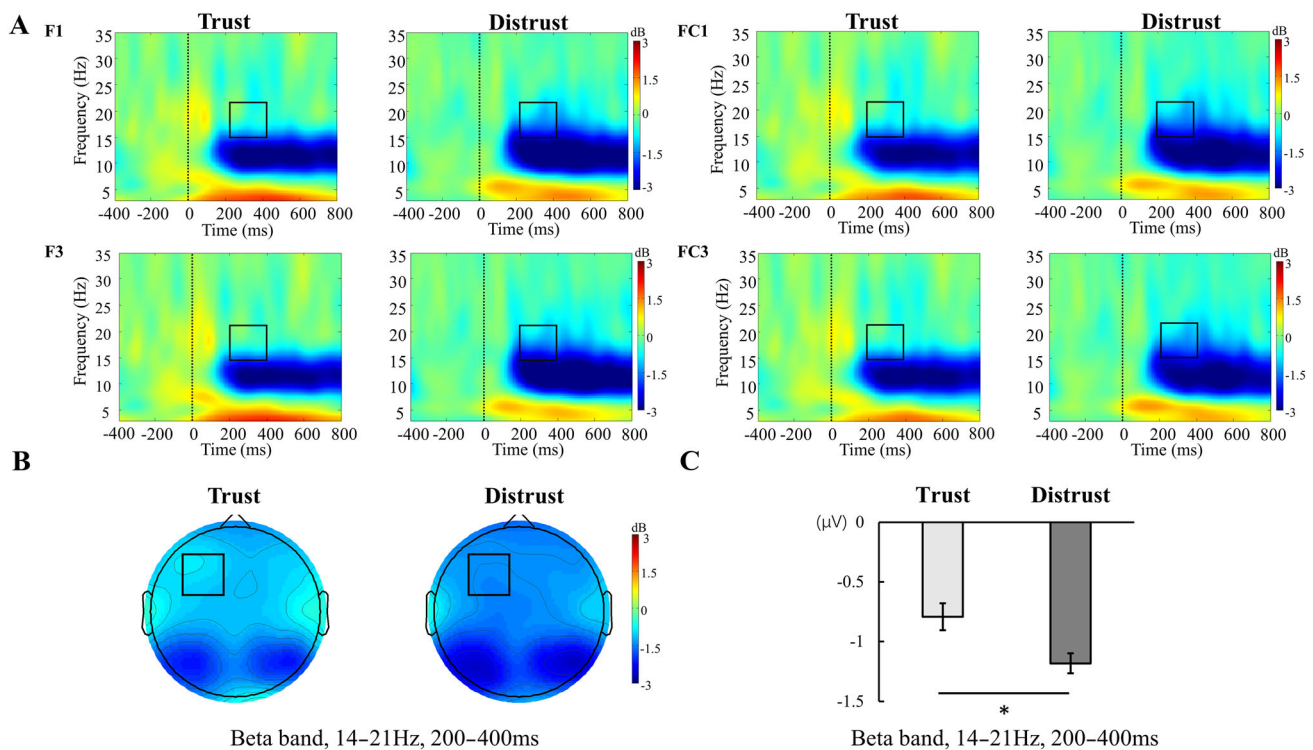
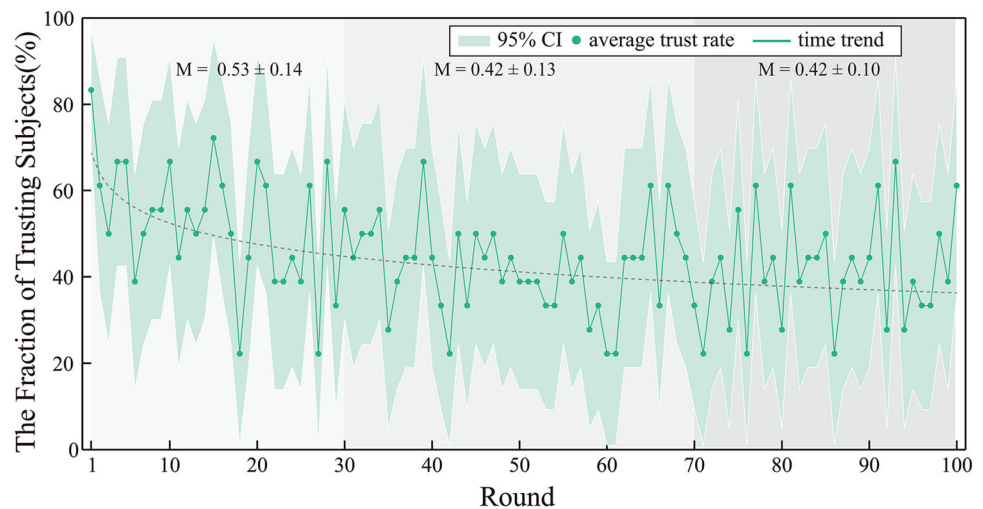


Fig. 4 Neurophysiological response differences during the decision generation stage at the item level. **A** Event-related spectral power (ERSP) at the F1, FC1, F3, and FC3 electrodes during the decision generation stage among the trust and distrust trials. Bold dark rectangles highlight the beta band (200–400 ms, 14–21 Hz) which statistically distinguished between the trust and distrust trials. The difference of the ERSP between the trust and distrust trials at $P < 0.05$

within the S-ROIs. The defined TF-ROIs are outlined in the rectangles. **B** The corresponding scalp topographies of the ERSP magnitudes among TF-ROIs (200–400 ms, 14–21 Hz). Bold dark rectangles include the S-ROIs (F1, F3, FC1, and FC3). **C** Mean ERSP magnitudes of the beta band spectral power within the defined S-ROIs and TF-ROIs in the trust game (14–21 Hz, 200–400 ms, F1, F3, FC1, and FC3, \pm SEM, $*P < 0.05$).

$P = 0.001$, $\eta_p^2 = 0.506$). The *post-hoc* tests revealed that the beta band oscillation was significantly stronger in the trust than in the distrust trials ($M_{\text{trust}} = -0.792 \pm 0.226$; $M_{\text{distrust}} = -1.181 \pm 0.193$, $P = 0.001$).

- Beta band power predicts trust choices at the individual level

After characterizing the neural indicators to discriminate trusting/distrusting choices, we asked whether the defined indicator could predict trusting at the individual level during the decision generation stage. This is an important question but it has rarely been reported that the neural activity during the decision generation

stage could predict a given individual trust choice. To answer this question, we measured the acquired beta band oscillation power and used the logit model to test whether the defined oscillatory power could significantly predict trusting using the HLM [39]. In this study, single-trial data were nested. The logit model was computed with the following equations:

$$\text{Trust choice}(1/0) = \beta_0 + \beta_1 \text{beta band} + \beta_2(\text{inter}_{\text{beta band} * \text{trust rates}})$$

In the equations, the trust choice in each trial was predicted by the intercept (β_0), the independent variable (β_1), and the interaction variable (β_2). The β_1 was the averaged beta band power of the pre-defined TF-ROIs (14–21 Hz) and S-ROIs (F1, F3, FC1, and FC3), and the β_2 was the interactions between the beta band power and trust rates.

The results of HLM showed that an individuals' trust choice was positively predicted by the beta oscillation power among the defined TF-ROIs and S-ROIs across all trials, as indicated by a significant regression: $n = 1649$, $b = 0.317$, $\text{SEM} = 0.033$, $t = 9.512$, $P < 0.001$, $\text{OR} = 1.372$, $95\% \text{CI} = [1.279, 1.473]$. The results suggested that the greater the beta band power, the more likely the trust choice at the individual level. Furthermore, we found that the trust rates also moderated the relationship between the trust choices and the beta band power: $n = 1649$, $b = -0.670$, $\text{SEM} = 0.051$, $t = -3.054$, $P < 0.001$, $\text{OR} = 0.512$, $95\% \text{CI} = [0.459, 0.571]$. The reverse moderating function of the trust rates on the prediction effects suggested that the lower the individual trust rates, the better the prediction effects of beta band power.

3. Validation analyses

The predictability of trust behaviors on the basis of neural activity of the frontocentral beta band oscillation (14–20 Hz, 200–400 ms, F1, F3, FC1, and FC3) during the trust generation stage was examined [37, 38]. Specifically, relative to chance prediction (50%), k-fold ($k = 10$) cross-validation suggested that the neural oscillation variables significantly and accurately predicted trust choices [sensitivity (one-tailed test): $t = 2.122$, $P = 0.031$; specificity: $t = 5.477$, $P < 0.001$; generalization rate: $t = 5.867$, $P < 0.001$].

4. Outcome feedback on trial_{n-1} moderates the prediction effect on trial_n

To determine whether the prediction effect was moderated by the outcome feedback on trial_{n-1}, we used the following equations:

$$\begin{aligned} \text{Trust choice}(1/0) = & \beta_0 + \beta_1(\text{beta band}) \\ & + \beta_2(\text{reciprocity feedback}) \\ & + \beta_3(\text{betrayal feedback}) \\ & + \beta_4(\text{inter}_{\text{beta band} * \text{reciprocity feedback}}) \\ & + \beta_5(\text{inter}_{\text{beta band} * \text{betrayal feedback}}) \end{aligned}$$

In the above equations, the trust choice in each trial was predicted by the intercept (β_0), independent variables ($\beta_1, \beta_2, \beta_3$), and interaction variables (β_4, β_5). The β_0 was centered and specified as a random variable. The β_1 was the averaged beta band power of pre-defined TF-ROIs (14–21 Hz, 200–400 ms) and S-ROIs (F1, F3, FC1, FC3). The outcome feedback on trial_{n-1} was defined as two dummy variables: “reciprocity feedback” = 1, else = 0 and “betrayal feedback” = 1, else = 0. The β_2 and β_3 was the variable of the reciprocity feedback and betrayal feedback on trial_{n-1} respectively. The β_4 and β_5 were the interactions between the beta band power on trial_n and the outcomes feedback on trial_{n-1}.

The HLM analysis showed that the interaction between the beta band power and the reciprocity feedback on trial_{n-1} was significant ($n = 1646$, $b = 0.272$, $\text{SEM} = 0.086$, $t = 3.160$, $P = 0.002$, $\text{OR} = 1.313$, $95\% \text{CI} = [1.109, 1.554]$; Fig. 5). However, there were no significant interaction between the beta band power and the betrayal feedback on trial_{n-1} ($b = 0.021$, $\text{SEM} = 0.025$, $t = 0.839$, $P = 0.413$, $\text{OR} = 1.021$, $95\% \text{CI} = [0.969, 1.077]$) and the distrust choice on trial_{n-1} ($b = -0.013$, $\text{SEM} = 0.019$, $t = -0.719$, $P = 0.482$, $\text{OR} = 0.987$, $95\% \text{CI} = [0.948, 1.026]$).

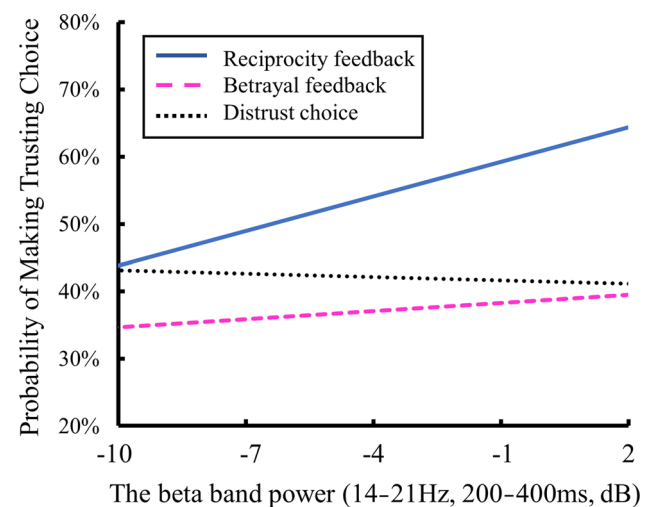


Fig. 5 The outcomes feedback on trial_{n-1} moderates the predictive effect of beta band power on individual trust choices on trial_n. The simple regression line is the regression line of the beta band power as the indicator on different conditions of the moderator. The horizontal axis represents the defined beta band oscillation power indicator, and the vertical axis represents the probability of making the trust choice in a single trial.

Discussion

Large-scale interpersonal trust pervades human social life. We asked whether there is a neural index that can predict an individual's trust choice in the trust scenario (i.e., decision generation stage) before a behavioral trust decision output? To address this issue, we conducted an iterative trust game and used an EEG approach to assess the neural oscillations during the decision generation stage to predict individuals' subsequent trust decisions.

The Tit-for-Tat Strategy is Adopted in the Trust Game

Regarding the behavioral results, the individuals' trust behaviors aligned with the trustees' reciprocity responses in this iterative trust game [26, 27]. Considering the behavioral trial-by-trial and time series analysis, the results were consistent with previous studies and suggested that the tit-for-tat strategy was adopted in the current study [40, 41]. Specifically, as expected, the participants were sensitive to the trial_{n-1} outcome feedback (i.e., reciprocity feedback, betrayal feedback, and distrust choices) in that they were more prone to trust after receiving direct positive reciprocity feedback and to distrust after receiving negative betrayal feedback (Fig. 2A). The above results revealed an adaptive and flexible strategy during ongoing exchanges (i.e., tit-for-tat strategy [28, 42]) so that individuals can benefit themselves, protect their own outcomes, and avoid exploitation [43].

More intriguingly, according to the results of RTs, we found that individuals were prone to make distrust choices. Specifically, the results showed longer RTs for the decision-making after receiving reciprocity feedback than that after betrayal feedback and distrust choices on trial_{n-1} (Fig. 2b). That is, individuals activate more inhibitory cognitive resources to solve a current cognitive conflict between the observed reciprocity behavior and the inherent distrust cognitive state in favor of a trust choice [44].

It is important to note that the behavioral results verified the credibility of the iterative trust game paradigm for detecting individual trust behaviors, consistent with previous studies [27, 29]. Specifically, the trial-by-trial results showed that trust choices were influenced by the trial_{n-1} outcome feedback, and the probability of trust choices on trial_n was significantly higher after receiving reciprocity feedback than that after receiving betrayal feedback on trial_{n-1} . Conversely, if the participants thought the opponent's behavior was random, the feedback from trial_{n-1} did not affect the trust choices of trial_n [18]. Accordingly, we suggest that the participants believed the opponent's

behaviors were not random and were in response to behaviors of his/her opponent.

Frontocentral Beta Band Oscillations Hierarchically Predict Trust Choices

During the iterative trust game, the convergent results supported our hypothesis and revealed that the frontocentral beta band power (14–21 Hz, 200–400 ms) in the decision generation stage was an important anticipatory indicator, as it reliably distinguished trust decision-making within an individual but also coded trust choices across different individuals. These conclusions extend findings from previous studies [45–47], and suggest that the cognitive inhibition-related beta band power can predict individuals' trust choices.

In particular, at the item level, the results during the decision generation stage were similar to those during the decision output stage, and suggested that there was greater cognitive inhibition-related beta band power in the trust condition than that in the distrust condition, consistent with previous studies [16, 48]. For example, Fu *et al.* adopted the identical iterative trust game task, and found that the trust decision exhibited higher frontocentral beta band power than the distrust decision, suggesting more cognitive inhibition is exerted on trusting choices [16]. At the individual level, we found that those with larger frontocentral beta band power were more likely to select trust choices in single trials, extending our conclusions. Combined with behavioral results, we confirmed that the distrust choice seems to be the advantageous option for participants, and individuals tend to make distrust choices to avoid exploitation by the trustee and obtain higher earnings. During the trust generation stage, once individuals prepare to make trust decisions, frontocentral beta band activity is required for the successful inhibitory control over this distrust intention, and this inhibitory mechanism may serve a long-term cooperative relationship in the iterative trust game [49, 50]. Furthermore, the prediction effects of beta band oscillations were moderated by the individual trust rates. In line with previous studies, we suggest that the process of cognitive inhibition is affected by individual differences in social trust [17, 18, 51]. For example, Declerck and colleagues demonstrated that the motivation to cooperate is influenced by cognitive inhibition-related prefrontal cortex, and the more pro-self the individual, the higher the activation of cognitive inhibition-related areas when performing pro-social behaviors during social exchange [25].

Outcome Feedback on Trial_{n-1} Moderates the Predictive Effects on Trial_n

Intriguingly, at the trial-by-trial level, the predictive effect of the beta band power during the predefined ROIs was moderated by the outcome feedback received on trial_{n-1}. Extending previous research [26, 29, 52], our results demonstrated that the associations between beta band power and trust choices were more pronounced after receiving trust reciprocity feedback than those after receiving trust betrayal feedback and distrust choice. This again suggests that the individuals' advantageous option is distrust, and once the reciprocity feedback is presented, the conflict between the observed reciprocity behavior and the inherent distrust cognitive intention arises. The results of the trial-by-trial moderated analysis implied that individuals prepared the trust choice during the trust generation stage in a flexible, selective manner by inhibiting an implicit advantageous option and by trust-shifting, which can conserve cognitive resources and ensure successful social interactions [15]. Moreover, we found a neural marker in the left, but not the right frontocentral regions. This finding may support Bernal and Altman's opinion that cognitive inhibition involves left-lateralized brain regions, while motor inhibition involves right-lateralized regions [53].

The Fast Neural Prediction Model

Based on the hierarchical but convergent electrophysiological results, we propose a fast neural prediction model for human trust behavior. In the social iterative trust situation, according to the decision circuit model [13], three distinct stages of trust behavior are interrelated. As illustrated in Fig. 6, the fast neural prediction model demonstrates that the cognitive inhibition-related beta band neural oscillations during the trust generation stage can forecast individuals' subsequent trust choices, and the predictive effects are moderated by the outcome feedback

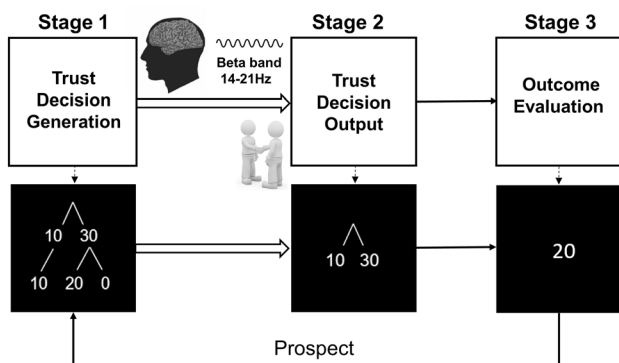


Fig. 6 The fast neural prediction model of the Trust Game.

of prior trials [54–57]. The predictive effect through the process of cognitive inhibition may help individuals adapt their task performance during the trust game [16, 26].

Although our study provides substantial new information about the neural basis of interpersonal trust, it is also limited in several important ways. First, we demonstrated that frontocentral beta band oscillations can predict trust behavior. Other important features, such as the resting state and interactions functional regions of the brain, need to be further explored to identify their role in predicting trust [58, 59]. In addition, to validate the predictive effect, additional samples can be used as the test set to test the predictive effect of the predictive model in future. Second, we used 50% reciprocity rates to avoid the influence of reciprocal behaviors, and future research could use a real interpersonal trust game to investigate trustors' strategic responses to their partners. Finally, existing research has suggested that gender differences and the menstrual cycle may also affect trust behaviors [43, 60]. Thus, a more complete understanding of the neural prediction mechanism could be achieved by taking into account these potential moderating variables in future research.

Conclusions

In summary, the present study provides novel and reliable neurophysiological insight into the cognitive processes of interpersonal trust by shedding new light on the predictive effect of beta band oscillations prior to trust behaviors. This study may not only help researchers understand the time course of neural activity underlying human trust decisions, but also have greater value by providing evidence from noninvasive EEG measurement.

Acknowledgements This work was supported by the National Natural Science Foundation of China (71942002 and 31771238) and the National Social Science Foundation of China (19ZDA361).

Conflict of Interests The authors declared no conflicts of interest with respect to the authorship or the publication of this article.

References

1. Balliet D, Van Lange PAM. Trust, conflict, and cooperation: A meta-analysis. *Psychol Bull* 2013, 139: 1090–1112.
2. Glaeser EL, Laibson DI, Scheinkman JA, Soutter CL. Measuring trust. *Q J Econ* 2000, 115: 811–846.
3. Haas BW, Ishak A, Anderson IW, Filkowski MM. The tendency to trust is reflected in human brain structure. *Neuroimage* 2015, 107: 175–181.
4. Carré JM, Baird-Rowe CD, Hariri AR. Testosterone responses to competition predict decreased trust ratings of emotionally neutral faces. *Psychoneuroendocrinology* 2014, 49: 79–83.

5. Hooper JJ, Sutherland CAM, Ewing L, Langdon R, Caruana N, Connaughton E *et al.* Should I trust you? autistic traits predict reduced appearance-based trust decisions. *Br J Psychol* 2019, 110: 617–634.
6. Biasiucci A, Franceschiello B, Murray MM. Electroencephalography. *Curr Biol* 2019, 29: 80–85.
7. Karamzadeh N, Medvedev A, Azari A, Gandjbakhche A, Najafzadeh L. Capturing dynamic patterns of task-based functional connectivity with EEG. *Neuroimage* 2013, 66: 311–317.
8. Buzsaki G, Logothetis N, Singer W. Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 2013, 80: 751–764.
9. Buzsaki G, Watson BO. Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues Clin Neurosci* 2012, 14: 345–367.
10. Gross J. Analytical methods and experimental approaches for electrophysiological studies of brain oscillations. *J Neurosci Methods* 2014, 228: 57–66.
11. Berg J, Dickhaut J, McCabe K. Trust, reciprocity, and social-history. *Games Econ Behav* 1995, 10: 122–142.
12. Johnson ND, Mislin AA. Trust games: A meta-analysis. *J Econ Psychol* 2011, 32: 865–889.
13. Platt ML. Neural correlates of decisions. *Curr Opin Neurobiol* 2002, 12: 141–148.
14. Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. *Cogn Process* 2007, 8: 159–166.
15. Thornton MA, Weaverdyck ME, Tamir DI. The social brain automatically predicts others' future mental states. *J Neurosci* 2019, 39: 140–148.
16. Fu C, Yao X, Yang X, Zheng L, Li J, Wang Y. Trust game database: behavioral and EEG data from two trust games. *Front Psychol* 2019, 10: 2656.
17. Wang Y, Jing Y, Zhang Z, Lin C, Valadez EA. How dispositional social risk-seeking promotes trusting strangers: evidence based on brain potentials and neural oscillations. *J Exp Psychol Gen* 2017, 146: 1150–1163.
18. Wang Y, Zhang Z, Jing Y, Valadez EA, Simons RF. How do we trust strangers? The neural correlates of decision making and outcome evaluation of generalized trust. *Soc Cogn Affect Neurosci* 2016, 11: 1666–1676.
19. Aimone JA, Houser D. What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Exp Econ* 2012, 15: 571–588.
20. Bohnet I, Zeckhauser R. Trust, risk and betrayal. *J Econ Behav Organ*. 2004, 55: 467–484.
21. Rand DG. Cooperation, Fast and Slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychol Sci* 2016, 27: 1192–1206.
22. Zeckhauser R, Bohnet I, Greig F, Herrmann B. Betrayal aversion: evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *Am Econ Rev* 2008, 98: 294–310.
23. Engel AK, Fries P. Beta-band oscillations – signaling the status quo?. *Curr Opin Neurobiol* 2010, 20: 156–165.
24. Huster RJ, Enriquez-Geppert S, Lavallee CF, Falkenstein M, Herrmann CS. Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. *Int J Psychophysiol* 2013, 87: 217–233.
25. Declerck CH, Boone C, Emonds G. When do people cooperate? The neuroeconomics of prosocial decision making. *Brain Cogn* 2013, 81: 95–117.
26. Delgado MR, Frank RH, Phelps EA. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 2005, 8: 1611–1618.
27. King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz S, Montague PR. Getting to know you: Reputation and trust in a two-person economic exchange. *Science* 2005, 308: 78–83.
28. Maurer C, Chambon V, Bourgeois-Gironde S, Leboyer M, Zalla T. The influence of prior reputation and reciprocity on dynamic trust-building in adults with and without autism spectrum disorder. *Cognition* 2017, 172: 1–10.
29. Fouragnan E, Chierchia G, Greiner S, Neveu R, Avesani P, Coricelli G. Reputational priors magnify striatal responses to violations of trust. *J Neurosci* 2013, 33: 3602–3611.
30. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007, 39: 175–191.
31. Thielmann I, Hilbig BE. Trust: An integrative review from a person-situation perspective. *Rev Gen Psychol* 2015, 19: 249–277.
32. Chen J, Zhong J, Zhang Y, Li P, Zhang A, Tan Q *et al.* Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neurosci Lett* 2012, 517: 65–70.
33. Pak R, McLaughlin AC, Bass B. A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. *Ergonomics* 2014, 57: 1277–1289.
34. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004, 134: 9–21.
35. Bigdely-Shamlo N, Mullen T, Kothe C, Su K, Robbins K. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform* 2015, 9: 16.
36. Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. *Trends Cogn Sci* 2004, 8: 204–210.
37. Knutson B, Rick S, Wimmer E, Prelec D, Loewenstein G. Neural predictors of purchases. *Neuron* 2007, 53: 147–156.
38. Price CJ, Ramsden S, Hope TMH, Friston KJ, Seghier ML. Predicting IQ change from brain structure: a cross-validation study. *Dev Cogn Neurosci* 2013, 5: 172–184.
39. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. 2 ed. 2002, Thousand Oaks, CA: Sage.
40. Engle-Warnick J, Slonim RL. The evolution of strategies in a repeated trust game. *J Econ Behav Organ* 2004, 55: 553–573.
41. Hoffman M, Yoeli E, Nowak MA. Cooperate without looking: why we care what people think and not just what they do. *Proc Natl Acad Sci U S A* 2015, 112: 1727–1732.
42. Koscik TR, Tranel D. The human amygdala is necessary for developing and expressing normal interpersonal trust. *Neuropsychologia* 2011, 49: 602–611.
43. Lemmers-Jansen ILJ, Krabbendam L, Veltman, DJ, Fett AKJ. Boys vs. girls: Gender differences in the neural development of trust and reciprocity depend on social context. *Dev Cogn Neurosci* 2017, 25: 235–245.
44. Badre D, Kayser AS, D'Esposito M. Frontal cortex and the discovery of abstract action rules. *Neuron* 2010, 66: 315–326.
45. Wolff A, Gomez-Pilar J, Nakao T, Northoff G. Interindividual neural differences in moral decision-making are mediated by alpha power and delta/theta phase coherence. *Sci Rep* 2019, 9: 4432.
46. Ai SZ, Yin Y, Chen Y, Wang C, Sun Y, Tang X *et al.* Promoting subjective preferences in simple economic choices during nap. *Elife* 2018, 7: e40583.
47. Hu L, Iannetti GD. Neural indicators of perceptual variability of pain across species. *Proc Natl Acad Sci U S A* 2019, 116: 1782–1791.
48. Wagner J, Makeig S, Gola M, Neuper C, Müller-Putz GR. Distinct β band oscillatory networks subserving motor and

- cognitive control during gait adaptation. *J Neurosci* 2016, 29: 12675–12685.
49. Axelrod R. Effective choice in the prisoner's dilemma. *J Conflict Resolut* 1980, 24: 3–25.
 50. Zhao K, Smillie LD. The role of interpersonal traits in social decision making: exploring sources of behavioral heterogeneity in economic games. *Pers Soc Psychol Rev* 2015, 19: 277–302.
 51. Bodmer B, Mueckschel M, Roessner V, Beste C. Neurophysiological variability masks differences in functional neuroanatomical networks and their effectiveness to modulate response inhibition between children and adults. *Brain Struct Funct* 2018, 223: 1797–1810.
 52. Fareri DS, Chang LJ, Delgado MR. Effects of direct social experience on trust decisions and neural reward circuitry. *Front Neurosci* 2012, 6: 148.
 53. Bernal B, Altman N. Neural networks of motor and cognitive inhibition are dissociated between brain hemispheres: an fMRI study. *Int J Neurosci* 2009, 119: 1848–1880.
 54. Bornstein AM, Khaw MW, Shohamy D, Daw ND. Reminders of past choices bias decisions for reward in humans. *Nat Commun* 2017, 8: 15958.
 55. Cisek P. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos Trans R Soc Lond B Biol Sci* 2007, 362: 1585–1599.
 56. Cisek P, Pastor-Bernier A. On the challenges and mechanisms of embodied decisions. *Philos Trans R Soc Lond B Biol Sci* 2014, 369: 315–318.
 57. Pennartz CMA, Ito R, Verschure PFMJ, Battaglia FP, Robbins TW. The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends Neurosci* 2011, 34: 548–559.
 58. Zhang YY, Xu L, Liang ZY, Wang K, Hou B, Zhou Y *et al*. Separate neural networks for gains and losses in intertemporal choice. *Neurosci Bull* 2018, 34: 725–735.
 59. Li J, Shen J, Liu S, Chauvel M, Yang W, Mei J *et al*. Responses of patients with disorders of consciousness to habit stimulation: a quantitative EEG study. *Neurosci Bull* 2018, 34: 691–699.
 60. Buser T. Digit ratios, the menstrual cycle and social preferences. *Games Econ Behav* 2012, 76: 457–470.