Cerebral Cortex, 2025, **35**, bhaf191
https://doi.org/10.1093/cercor/bhaf191
Advance access publication date 21 July 2025
Original Article

Neurocomputational mechanisms underlying how social status affects learning of trust behavior

Siying Li[1], Jean-Claude Dreher[2,3], Edmund Derrington[2,3], Haoke Li[4], and Chen Qu[5,*]

[1]Faculty of Education, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin 130024, China
[2]Laboratory of Neuroeconomics, Institut des Sciences Cognitives Marc Jeannerod, CNRS, 67 Bd Pinel, 69675 Bron, Lyon, France
[3]Université Claude Bernard Lyon 1, Avenue Jean Capelle, 69100 Villeurbanne, Lyon, France
[4]School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, 2001 Longxiang Avenue, Longgang District, Shenzhen, Guangdong 518172, China
[5]Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, School of Psychology, Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, No. 55, West Zhongshan Avenue, Tianhe District, Guangzhou, Guangdong 510631, China
[*]Corresponding author: School of Psychology, South China Normal University, No.55, West Zhongshan Avenue, Tianhe District, Guangzhou, Guangdong 510631, China. Email: fondest@163.com

Social status, as a prominent social characteristic, exerts a significant influence on various aspects of life. However, there is only limited behavioral and neural evidence regarding the relationship between social status and the construction of trust. In this study, we used computational modeling and functional magnetic resonance imaging to unveil the trajectory of trust-related processing by using a repeated trust game. Human participants assumed the role of trustor and engaged in interactions with fictitious partners (trustees) who varied in social status. Participants were more inclined to trust Superiors than Inferiors and gradually modified their trust decisions based on their partners' reciprocity. Furthermore, we unveiled the neurocomputational mechanisms of two cognitive processes: (i) prior-based static modulation supported by the ventromedial prefrontal cortex (vmPFC), amygdala, and their neural coupling, and (ii) the reward network engaged in feedback-based dynamic modulation. We also found that prior bias in the social value of social status can reduce the reliance on the feedback-based dynamic modulation rooted in the vmPFC and ventral striatum. The present findings enhance the understanding of the neural representations of how social status modulates trust-related processing and trustworthiness updating.

Keywords: fMRIsocial decision-making; social learning; social status; trust.

Introduction

Trust is a fundamental element in fostering cooperation and reciprocity, and interhierarchical trust has long been considered to facilitate coordination on group tasks and highly impact the efficiency of societal functioning (Magee and Galinsky 2008; Cheng et al. 2013). Interpersonal trust constitutes a social dilemma, characterized by inherent uncertainty, which arises from the potential for greater collaborative gains. In this context, trustors willingly expose their own interests to others, despite the risk of betrayal (FeldmanHall and Shenhav 2019; Krueger and Meyer-Lindenberg 2019; Bellucci and Dreher 2022). To mitigate uncertainty, individuals assess the trustworthiness of others and optimize decisions based on two types of information that influence each other mutually: prior information and direct experience (Tzieropoulos 2013; Yu et al. 2014; Bellucci et al. 2017, 2019). Social status, as effective prior information, can influence trust-building (Blue et al. 2018, 2020; Foncelle et al. 2022; Li et al. 2023). More specifically, we have shown previously in a repeated trust game that a high-status partner (trustee, ie receiving trust) gains more trust than a low-status partner. This effect, called the superior bias, may be due to the fact that a higher social status provides higher social value. This bias may explain why people judge high-status wrongdoers less harshly than low-status individuals who commit the same types of acts (Bowles and Gelfand, 2009). Here, we asked two questions related to the integration of social status information into trust decisions. First, how is social status represented by neural mechanisms when solving trust-related issues? Second, how does social status influence trustworthiness learning and updating? Driven by these questions, we conducted a functional magnetic resonance imaging (fMRI) experiment using computational modeling, to characterize the neural activation patterns related to trust-related decisions colored by social status.

The influence of social status on trust-building resembles prior-based top–down processing, linked by their intersecting characteristics, ie the attributes of value and sociality. For the attribute of value, trust reciprocity is generated by the motivation of pursuing mutual interests that engage key regions anchored in the reward network including the ventromedial prefrontal cortex (vmPFC) and ventral striatum (VS) (Knutson 2005; Bartra et al. 2013; Ruff and Fehr 2014; Lieberman et al. 2019; Yang et al. 2019). Social status indicates a consensual implication of

Received: October 31, 2024. **Revised:** April 9, 2025. **Accepted:** June 24, 2025
© The Author(s) 2025. Published by Oxford University Press. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

relative value that each member contributes to the group (Berger et al. 1980; Magee and Galinsky 2008; Polman et al. 2013). The implication of social value in social status may serve as the underlying motivation for its influence on a wide range of social interactions by recruiting reward processing (Li et al. 2021). For the attribute of sociality, social status could modulate brain regions (such as the amygdala) that are involved in social functions, thereby integrating social status information into the evaluations of trustworthiness. In trust-related processing, the amygdala maintains an individual's appropriate assessments of social functioning with respect to trust (Adolphs et al. 1998; Amaral 2006; Baumgartner et al. 2009; van Honk et al. 2013; Sladky et al. 2021). In face perception, the amygdala plays an important role in evaluating facial trustworthiness (Engell et al. 2007; Todorov and Duchaine 2008; Todorov and Engell 2008) and learning unfamiliar others' trustworthiness through stimulus generalization based on facial features (Feldmanhall et al. 2018). In general, the intersecting characteristics promote the interaction of social status and trust.

Repeated trust-related interactions with prior information on social status constitute a reinforcement learning process in a social context (Fareri et al. 2012, 2015; Fouragnan et al. 2013; Tzieropoulos 2013; Krueger and Meyer-Lindenberg 2019). The role of reinforcement learning in repeated trust behaviors has been extensively explored in a substantial body of research (Fareri et al. 2012, 2015; Ligneul et al. 2016, 2017; Bellucci et al. 2019; Krueger and Meyer-Lindenberg 2019; Janet et al. 2022). Specifically, repeated trust-related interactions can be treated as a feedback-based dynamic cognitive updating process in which individuals update their expectations of partners' reciprocity from outcomes under the incentive of monetary benefit or potential social reward (Krueger et al. 2007; Delgado et al. 2023). During this process, the VS and vmPFC encode prediction error (PE) signals of reciprocity by comparing the expected and actual outcomes (Joiner et al. 2017; Lockwood and Klein-Flügge 2020). Since social status is a high-level social concept, there are two possibilities for its interaction with trial-and-error learning. First, social status may promote part of learning in which the outcome aligns with the prior expectation, leading to asymmetric learning. Alternatively, if decisions are guided by a top–down cognitive process, this could weaken feedback-based bottom–up learning and result in individuals becoming less reliant on the outcome of direct interactions. In the current study, we employed a computational modeling approach in the context of a repeated trust game to formalize and test hypotheses regarding the integration of social status information into trust assessments.

In this study, combined with computational modeling and the technique of fMRI, we examined the impact of trustees' social status on trust decisions. At the behavioral level, with reinforcement learning models, we decoded the way that social status integrates into trustworthiness learning and revealed the cognitive factors that social status works on. At the neural level, we tested the prior-based impact on cognitive processing by comparing the neural activation in different levels of social status conditions. In further investigation of the interaction between prior-based and feedback-based modulation, we explored the effect of social status on feedback-based social learning processing.

Materials and methods

Participants

Twenty-eight right-handed participants were involved in this experiment (17 females;11 males; age M ± SD: 21.25 ± 2.99). All participants were recruited via an online advertisement. All participants gave informed written consent, reported basic information (such as age, gender, and handedness), and underwent screening for fMRI experiment contraindications before the experiment. None of them had a history of neurological or psychiatric disorders. They were paid 100 yuan for participation and gained extra monetary rewards (0 to 7.5 yuan) from the trust game. Three males and one female were excluded from data analyses because of excessive head movement during the fMRI scanning. In the final analyses, 24 participants were included (16 females; 8 males; age M ± SD: 20.71 ± 1.83). The study involving human participants was reviewed and approved by the Ethics Review Committee of South China Normal University (Approval Number: SCNU-PSY-2019-3-033).

Social status induction

The social status of trustees was manipulated as relative ranks of comprehensive capacities including income, occupational prestige, and education (Adler et al. 2000; Piff et al. 2010; Kraus et al. 2011), presented in a cover story and assigned using a star system (Li et al. 2023). Before the trust game, a fictitious experimental project named the Decision Information Collection Project was introduced to participants to produce a sense of immersion. Participants were informed that the Decision Information Collection project solicited subjects' social decisions and formed a database of over 500 individuals. The experimenter clarified that participation in the Decision Information Collection Project was voluntary and unrelated to the current experiment. If they volunteered for this project, they would become someone's partner in another experiment (ie the decisions that they made in this task would be presented as data when they were paired with another participant), and they were required to provide a photo for this experimental program. To ensure that participants believed in the authenticity of this project, if they decided to participate in the project, they would receive 5 yuan as compensation.

Following these instructions, the rules of the trust game and the information about their partners were introduced to the participants. They were informed that they would be Player A (i.e., trustors) and would interact with participants from earlier sessions of the Decision Information Collection project who played as Player B (ie trustee = person who receives trust). In addition, participants were told that their partners had been tested by questionnaires, which assessed their capabilities and basic information, such as their highest educational degree, income level, college entrance exam score, and occupation, implying their comprehensive capacity. Based on their relative superiority, the partners were ranked as Superior (marked as three stars), Intermediate (marked as two stars), or Inferior (marked as one star). This information regarding their partners' social statuses was shown to participants in the trust game.

Trust game

After the procedure of social status induction, participants were shown the four same-gender photos of trustees that were marked with stars to indicate their relative social status. They were informed that these four trustees, randomly selected from the Decision Information Collection project, would be their partners in the subsequent trust game. To mitigate the potential impact of gender, male participants interacted with four male trustees, and female participants interacted with four female trustees. Two of these trustees were manipulated as high status, represented by three stars, while the remaining two were assigned a low status, signified by a single star (Fig. 1A). Then, participants were asked

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

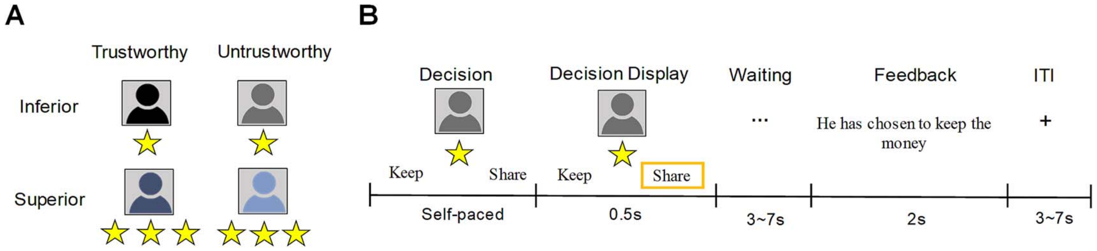

Fig. 1. Task schematic. A) Partners in experiment. In the experiment, participants engaged in a trust game with four partners; two of them were manipulated as high status, represented by three stars, while the remaining two were assigned a low status, signified by a single star. Unbeknownst to participants, partners had preprogrammed reciprocity rates such that: One high- and one low-status trustee had a high reciprocity rate (70%; trustworthy), and the other high- and low-status trustees had a low reciprocity rate (30%; untrustworthy). Photos of partners were taken from the Chicago Face database (Ma et al. 2015). The matching of facial stimuli, social status, and reciprocity rates between participants was balanced to exclude the potential effects of facial features and interactions among facial features, social status, and reciprocity rate. B) The trial procedure of the trust game. Participants completed a multishot binary version of the trust game. They were required to perform a binomial forced-choice task and decide whether to invest money in four same-gender trustees. In the decision phase (self-spaced), participants were endowed with 5 yuan and chose whether to keep the money for themselves or to share it with their partner. A decision made by a trustor to share money was described as an investment, making the trustee triple their money to 15 yuan on a given trial. Then, if the trustee decided to reciprocate, both players acquired half of the 15 yuan separately, equating to 7.5 yuan each. Conversely, if the trustee decided to keep all 15 yuan, the trustor obtained nothing. After participants made their decisions (displayed on the screen for 0.5 s) and waited a while (3 ~ 7 s), they were presented with feedback for 2 s.

to assess the trustworthiness of each trustee on a 9-point Likert scale (1 = not at all, 9 = a lot).

Following the first trustworthiness rating, participants completed the main task in the scanner. In the two widely applied multishot versions of the trust game, ie the binary-choice version and the continuous-choice version (Delgado et al. 2005; Fareri et al. 2012, 2015), we applied the binary-choice version trust game as the task for this experiment. Compared to the continuous-choice version, the binary-choice version offers a more straightforward approach for qualitatively differentiating the participants' choices as either trust or distrust. In this experiment, participants assumed the role of trustors. Unlike the continuous-choice version of the trust game, the participants did not send a specific amount of money to trustees (for example, participants could send to the trustee any amount between 1 and 10 euros) (Gjoneska et al. 2019). They were required to perform a binomial forced-choice task and to decide whether to invest money in four same-gender trustees or not (Fig. 1B). Photos and social status information of the trustees were shown to participants. The initial endowment of the four trustees was equal, specifically 0 yuan. Participants were endowed with 5 yuan on each trial. The duration of this decision phase lasts until the participants complete their responses (self-paced). A decision made by a trustor to share money was described as an investment, making the trustee triple their money to 15 yuan on a given trial. Then, if the trustee decided to reciprocate, both players would acquire half of the 15 yuan separately, equating to 7.5 yuan each. Conversely, if the trustee decided to keep all 15 yuan, the trustor would obtain nothing on that trial. Another choice available to the participants was to keep the money, which let participants obtain 5 yuan with no money for the trustees on that trial. The participant's decision to keep or share money was displayed (decision display phase) on the screen for 0.5 s, which was followed by an interstimulus interval showing a jittered waiting phase for 3 ~ 7 s (ie randomized between 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, and 7 s). Then they were presented with one of three possible feedbacks (feedback phase) for 2 s based on their responses: "You have kept the money", "She/He has chosen to keep the money", or "She/He has chosen to share the money." Participants were informed before the experiment that one trial would be selected randomly at the end of the game, and the outcome of this trial, ie obtaining 0, 5, or 7.5 yuan, would serve as their reward. Each trial ended with an intertrial interval (ITI) showing another jittered fixation cross for 3 ~ 7 s (randomized between 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, and 7 s).

Unbeknownst to participants, trustees had preprogrammed reciprocity rates such that: one high- and one low-status trustee had a high reciprocity rate (70%), and the other high- and low-status trustees had a low reciprocity rate (30%). The trust game consisted of 96 trials, evenly distributed across eight blocks of 12 trials each. In each block, the participants interacted with each trustee for 3 trials, and they played with these four trustees in random order in each block. In total, 24 trials per trustee condition were randomly administered across these blocks.

Eight neutral facial stimuli (four female facial stimuli and four male facial stimuli) were used for this task. These photos were taken from the Chicago Face database (Ma et al. 2015). Matching of facial stimuli, social status, and reciprocity rates between participants was balanced to exclude the potential effects of facial features and interactions among facial features, social status, and reciprocity rates. The presentation of the task and recording of behavioral responses were performed using E-Prime 2.0. After the trust game, participants rated the trustworthiness of the four trustees again outside of the scanner.

Data acquisition

Scanning was conducted on a Siemens 3-tesla Trio Scanner. Functional data were acquired using echo-planar imaging sequences with the following parameters: repetition time = 2,000 ms, echo time = 30 ms, flip angle = 90°, field of view = 224 mm, slice thickness = 3.5 mm. A total of 62 axial slices were acquired in interleaved order (in-plane resolution 2 × 2 mm). After functional scanning, high-resolution T1-weighted anatomical images (generalized autocalibrating partially parallel acquisitions, GRAPPA, 0.5 × 0.5 × 1 mm) were also acquired for each subject. The presentation of the task and recording of behavioral responses were performed using E-Prime software, version 2.0.

fMRI analysis

Neuroimage preprocessing

For fMRI data preprocessing, all images were preprocessed using SPM12 by the following steps: In general, functional images were slice-timing corrected, realigned, coregistered to the anatomical scan, normalized to Montreal Neurological Institute (MNI) space, and smoothed with an 8 mm isotropic Gaussian kernel. For further details, a six-parameter rigid-body transformation was used

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

for realignment correction. Each subject's structural image was coregistered to the average of the motion-corrected images using a 12-parameter affine transformation. Movement outliers were identified and excluded if head movements/translations were above 3 mm/rad. Four participants who met these criteria were excluded from the final analyses. The results of neural images were visualized using the MRIcroGL.

First-level design matrix

GLM1 was constructed to explore the effect of social status on trust decisions by modeling the decision to keep or share with each trustee, which consisted of eight regressors. Phases of decision display and feedback were modeled as two regressors of no interest separately. The other two GLMs that focus on the feedback phase were defined to investigate the learning process in the trust game. GLM2 modeled the decision phase, decision display, and feedback phase as three regressors and one regressor coding for model-derived parameters reflecting the PE. The decision phase and decision display phase were modeled as two regressors of no interest separately. GLM3 examined the role of social status on the different types of feedback, which included 12 regressors associated with the three types of feedback phase (trustee keep, trustee share, or trustor keep) for four trustees. The decision phase and decision display phase were modeled as two regressors of no interest separately. In these three GLMs, participant-specific movement parameters were included as regressors of no interest. A high-pass filter with a cutoff of 128 s was employed. Temporal autocorrelation was modeled using an AR(1) process.

Statistical inference

In the second-level analyses, participant-specific linear contrasts of these regressors were entered into a series of one-sample *t*-tests. For whole-brain analysis, activation was considered significant if it survived at the threshold of $P < 0.001$ uncorrected, and additionally survived after family-wise error (FWE) correction at the cluster level (FWEc $P < 0.05$ corrected). To provide more detailed information, we also reported results at an uncorrected voxel-wise threshold of $P < 0.001$ with a voxel-wise threshold of $P < 0.05$ after FWE correction in Table 1. In addition, small volume correction (SVC) with cluster-level FWE corrected *P*-values ($P < 0.05$) was used on a priori regions of interest (ROIs), including amygdala, vmPFC, and VS. For the SVC procedure, we used amygdala atlases from the automated anatomical atlas (aal) template and we used a priori ROIs from previous meta-analyses for vmPFC (MNI coordinate, $x = -2$, $y = 50$, $z = -16$, spheres of 12 mm) and VS (MNI coordinate, $x = \pm 16$, $y = 4$, $z = -14$, spheres of 6 mm) where activities in the region were correlated with reward valence and processing stages (Liu et al. 2011).

Psychophysiological interaction analysis

A psychophysiological interaction (PPI) analysis was conducted to test the presence of functional coupling between different brain regions in different conditions of the social status context. We performed a GLM that contained regressors that reflected the physiological effect, the psychological contrast of interest, and the psychophysiological interaction term. The procedures of PPI analysis were as follows: (i) establish the physiological effect, ie the time series of activity in the vmPFC seed regions in this study. Specifically, the time series for the peak voxel in the vmPFC was extracted by defining the VOI with a 6 mm sphere centered on the peak coordinate (left peak: $x = -8$, $y = 32$, $z = -10$; right peak: $x = 12$, $y = 34$, $z = -10$) which was identified in the group analysis for activation in the decision phase, collapsed across both high and low social status conditions; (ii) identify the psychological contrast of interest, where high status > low status in the decision phase; and (iii) create a PPI variable based on the above types, ie that forms the interaction term of source signal × experimental condition. With these variables, the effect of the psychophysiological interaction term was assessed for each participant and entered into a group-level analysis.

Model construction and estimation

In the trust game, participants learn the trustworthiness of different trustees through the reinforcement of feedback, which has the basic characteristics of reinforcement learning. Thus, four possible frameworks of social status integrated into trust construction were formalized and tested based on reinforcement learning with different hypotheses. The basic framework of these models includes the information updating phase and the decision phase. Using the Rescorla–Wagner updating rule, the updating phase decreases the difference between outcome and expectation to guide decision-making. The pursuit of benefit maximization, represented by expected value, drives the decision phase. We specifically exemplified a model with social value items and learning rates (SV&LR Model), which quantifies the impact of social status both on the updating phase and decision phase. For a given trustee context, the expectation of future reciprocation (P_{t+1}, Eq. (1)) was determined by the current expectations (P_t) and their discrepancy from the actual outcome (γ_t, $\gamma = 1$ when the trustee reciprocated, $\gamma = 0$ when the trustee keeps), referred to as the PE (PE_t, Eq. (2)). PE quantifies the updating phase and is further scaled by the learning rate (α), which is bounded between 0 and 1. For each trustee, the expectation of reciprocation was transformed into its expected value with a free parameter (θ) (Eq. (3)). The θ tested the role of social value that was accompanied by social status, which is bounded between −7.5 and 7.5. To enhance the perceptibility of the social value term, we have situated it within this framework that is consistent with conventional monetary value, given that 7.5 represents the maximum amount accessible to participants in a single trial. The social value of social status is a subjective value that varies among participants. Subjective values are also a source of motivation to drive decisions (Yu et al. 2021). Thus, we supposed that the difference in the social value term between Superior and Inferior may correlate with the participants' explicit behavioral bias when they processed the trust issue in trials involving the two different social statuses. Subsequently, to determine the likelihood of a participant sharing money with a specific partner [IP, Eq. (4)], the expected value [$EV(s)_t$] was computed using a softmax function [Eq. (5)]. The parameter β indicates whether a participant was inclined toward exploratory or exploitative behavior. See Supplementary data for details of the other three models.

To choose a more representative model, we used the Akaike Information Criterion (AIC; Akaike 1974), which applied a penalty scaled by the number of free parameters of a complicated model, as a metric of model fit and compared model fits using paired *t*-tests.

SV&LR Model:

$$P_{t+1} = P_t + \alpha_i * \mathbf{PE}_t \tag{1}$$

$$\mathbf{PE}_t = \gamma_t - P_t \tag{2}$$

$$\mathrm{EV(S)}_t = P_t * 7.5 + \theta_i \tag{3}$$

$$\mathrm{IP} = \frac{e^{\frac{\mathrm{EV(S)}_t}{\beta}}}{e^{\frac{\mathrm{EV(S)}_t}{\beta}} + e^{\frac{\mathrm{EV(K)}_t}{\beta}}} \tag{4}$$

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

Table 1. Results of whole-brain analysis.

Brain region	Hemisphere	Cluster size	BA	MNI			t-value
				x	y	z	
superior_share > superior_keep							
Cingulate gyrus[a]	R	144	31	20	−50	28	8.65
Medial frontal gyrus	B	208	11/32	−8	32	−10	5.78
				12	34	−10	4.63
Prediction error							
Medial frontal gyrus/caudate/putamen[a]	B	2,309	32/11	10	26	8	9.96
				−10	8	−6	7.44
				6	34	−14	7.19
Middle frontal gyrus[a]	L	452	8/9	−24	32	42	6.63
Parietal lobe/occipital lobe	B	8,867	7/19	8	−66	−6	6.33
Cerebellum	R	239		14	−48	−50	6.12
Cerebellum	R	248		40	−68	−36	5.85
Frontal lobe/superior frontal gyrus	R	275	8/9	18	42	46	5.72
Middle temporal gyrus	L	697	39	−46	−70	28	5.60

Note: All results survived at an uncorrected voxel-wise threshold of $P < 0.001$ with a cluster-wise threshold of $P < 0.05$ after FWE correction. R, right; L, left; B, bilateral; BA, Brodmann area. [a]indicates results also survived at an uncorrected voxel-wise threshold of $P < 0.001$ with a voxel-wise threshold of $P < 0.05$ after FWE correction.

$$\mathrm{LLE} = \sum_{t=1}^{n} \log\left(\mathrm{IP}, j_t\right) \qquad (5)$$

Results

Behavioral results

Trust game

The behavioral measures of interest were whether participants' trust-related decisions would vary as a function of the social status of partners and how they changed along with trustworthiness updating in trial-and-error learning (Fig. 2A). All statistical tests were two-sided. A Three-way repeated-measures ANOVA was conducted on the share rate with social status (Superior vs. Inferior), reciprocity rate (high vs. low), and Block (from Block1 to Block8) as within-subject variables. The observed significant main effect of social status [$F_{(1,23)} = 6.945$, $P = 0.015$, $\eta p^2 = 0.232$] and the main effect of reciprocity rate [$F_{(1,23)} = 23.762$, $P < 0.001$, $\eta p^2 = 0.508$] revealed that participants demonstrated a stronger propensity to share with Superiors ($M = 0.67$, $SD = 0.13$) than Inferiors ($M = 0.55$, $SD = 0.21$) and with partners who showed a high reciprocity rate ($M = 0.73$, $SD = 0.17$) than a low reciprocity rate ($M = 0.49$, $SD = 0.20$). Furthermore, we found the learning effect was characterized by the significant interaction of social status × Block [$F_{(7,161)} = 2.210$, $P = 0.036$, $\eta p^2 = 0.088$] and interaction of reciprocity rate × Block [$F_{(7,161)} = 6.054$, $P < 0.001$, $\eta p^2 = 0.208$]. Post hoc tests showed that the difference between the share rate of Superior and Inferior was more robust in the early stages [Block1: $t_{(23)} = -\ 4.112$, $P_{(\mathrm{Tukey})} = 0.008$] and then started to fade. However, post hoc tests revealed that the difference in share rate between trustworthy partners and untrustworthy partners became more pronounced at later stages [Block5: $t_{(23)} = 5.140$, $P_{(\mathrm{Tukey})} < 0.001$; Block6: $t_{(23)} = 4.246$, $P_{\ (\mathrm{Tukey})} = 0.006$; Block7: $t_{(23)} = 4.356$, $P_{\ (\mathrm{Tukey})} = 0.004$; Block8: $t_{(23)} = 5.434$, $P_{(\mathrm{Tukey})} < 0.001$). The three-way interaction of social status × reciprocity rate × Block was not significant [$F_{(7,161)} = 1.62$, $P = 0.13$, $\eta p^2 = 0.07$]. No other significant effect was found ($Ps > 0.05$).

One-way repeated-measures ANOVA was conducted to probe the effect of social status on reaction time. No significant difference was found ($Ps > 0.05$).

Subjective rating

Participants rated the trustworthiness of the trustees twice as a manipulation check (Fig. 2B). A three-way repeated-measures ANOVA was conducted on the trustworthy rating with social status (Superior vs. Inferior), reciprocity rate (high vs. low), and rating sequence (prerating and postrating) as within-subject variables. Results revealed a significant main effect of social status [$F_{(1,23)} = 18.344$, $P < 0.001$, $\eta p^2 = 0.444$], indicating that participants rate Superiors ($M = 5.85$, $SD = 0.92$) as more trustworthy than Inferiors ($M = 4.69$, $SD = 1.03$). The observed significant main effect of reciprocity rate [$F_{(1,23)} = 27.243$, $P < 0.001$, $\eta p^2 = 0.542$) revealed that participants rated partners who showed a high reciprocity rate ($M = 5.88$, $SD = 0.88$) as more trustworthy than those with low reciprocity rate ($M = 4.67$, $SD = 0.94$). We found a significant interaction of social status × rating sequence [$F_{(1,23)} = 4.738$, $P = 0.040$, $\eta p^2 = 0.171$]. The post hoc test demonstrated that the difference between Superior and Inferior was significant in prerating [$t_{(23)} = 4.793$, $P_{(\mathrm{Tukey})} < 0.001$], but not in postrating [$t_{(23)} = 2.561$, $P_{(\mathrm{Tukey})} = 0.067$], which indicates that the influence of social status had faded as participants learned trustees' reciprocity rates in the trust game. Results also revealed a significant interaction of reciprocity rate × rating sequence [$F_{(1,23)} = 26.590$, $P < 0.001$, $\eta p^2 = 0.536$]. The post hoc test demonstrated that the difference in rating score between partners with a high reciprocity rate and partners with a low reciprocity rate was significant in postrating [$t_{(23)} = 7.335$, $P_{(\mathrm{Tukey})} < 0.001$], but not in prerating [$t_{(23)} = -0.063$, $P_{(\mathrm{Tukey})} = 1.000$], indicating participants learned partners' reciprocity rates in the trust game. The three-way interaction of social status × reciprocity rate × test sequence was not significant [$F_{(1,23)} = 0.484$, $P = 0.493$, $\eta p^2 = 0.021$]. No other significant effect was found ($Ps > 0.05$).

Model comparisons and validation

The results of model estimation and comparison are shown in Supplementary Table S1. Based on AICs, we found that the SV&LR Model fitted the participants' data better than LR Model [$t_{(23)} = 5.256$, $P < 0.001$, Cohen's $d = 1.072$], LG Model [$t_{(23)} = 2.544$, $P = 0.018$, Cohen's $d = 0.519$], and SV&LG Model [$t_{(23)} = 3.209$, $P = 0.004$, Cohen's $d = 0.655$].

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

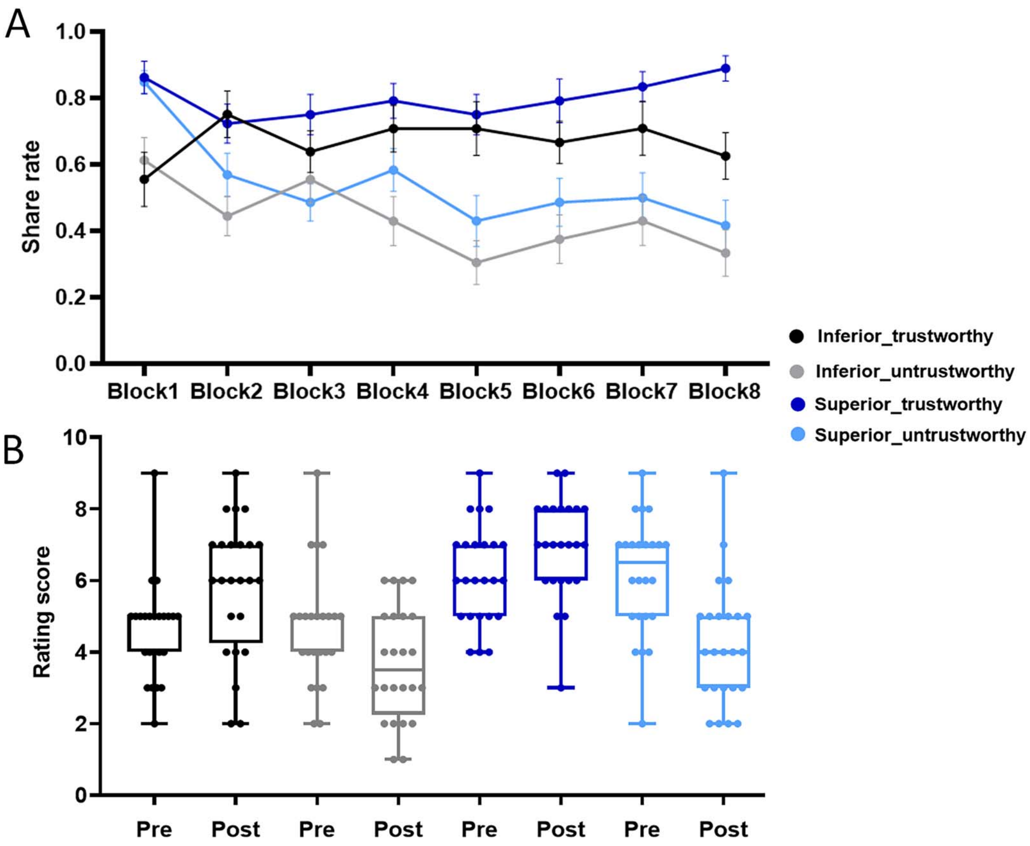

Fig. 2. Share rate in the trust game and trustworthiness rating. A) Mean percentage of trials in which participants shared in each block across the experiment condition of their trustees. B) Subjective trustworthiness rating before (pre) and after (post) the trust game. Error bars refer to SEM.

The above analysis suggests that the SV&LR Model is highly fit for capturing the characteristics of actual behavior (Supplementary material). Based on the SV&LR Model, we tested the relationship between social value and behavioral superior bias. The social value item as a subject-specific parameter measures the significance of various social statuses in individuals' perceptions. We found a significant positive correlation between the difference in social value term (i.e., social value $_{\text{Superior}}$ − social value $_{\text{Inferior}}$) and behavioral bias (ie share rate $_{\text{Superior}}$ − share rate $_{\text{Inferior}}$; $r=0.457$, $P=0.025$). This indicates that the higher the social value of social status, the higher the Superior bias in trust decisions from participants (trustors). We also found a significant negative correlation between the difference in learning rate (learning rate $_{\text{Superior}}$ − learning $_{\text{Inferior}}$) and behavioral bias ($r=-0.460$, $P=0.024$). This finding implies that the updating process influences the participants' share rate in the trust game.

Neuroimaging results

Decision phase

At the time of choice, participants evaluated the benefits of investing money versus keeping money. We mainly focus on whether and how decision-making is modulated by the social status of the trustee. Firstly, we conducted a whole-brain ANOVA of social status × decision, during the decision phase, to analyze their interaction (Supplementary material; Table S2). However, no significant results were found. Then, we conducted an exploratory analysis of the one-sample *t*-tests. The result revealed activation in vmPFC when contrasting share and keep decisions in the context of interacting with Superiors [left peak: *x*, *y*, $z=-8$, 32, −10, $t_{(23)}=5.78$, $Z=4.50$, FWEc $P=0.010$; right peak: *x*, *y*, $z=12$, 34, −10, $t_{(23)}=4.63$, $Z=3.85$, FWEc $P=0.010$] but not found in the partner context of Inferiors (Fig. 3A; Table 1). Activation of the vmPFC was stronger when sharing money with Superiors than when keeping money. To illustrate how the blood oxygen level-dependent (BOLD) signal varied with social status in the decision phase, we obtained parameter estimates from these regions (6 mm radius spheres with center at the reported peak coordinates) as shown in Fig. 3A.

This result prompted us to further explore functional connectivity since vmPFC has previously been shown to be functionally connected with brain regions associated with social cognition during value computations (Hare et al. 2010; Bellucci et al. 2019). We supposed that the functional coupling of the signal in vmPFC may vary when evaluating the trustworthiness of different partners. To verify this potential neural integration, we performed a PPI analysis using the vmPFC as seed regions. This analysis of social status will enhance our comprehension of the impact such activity has on subsequent behavior or neural activation. Results revealed that the left vmPFC was more strongly coupled to the right amygdala [*x*, *y*, $z=22$, 0, −14, $t_{(23)}=3.57$, $Z=3.15$, SVC $P=0.031$], and the right vmPFC was more strongly coupled to the bilateral amygdala [right amygdala: *x*, *y*, $z=22$, 0, −14, $t_{(23)}=3.86$, $Z=3.36$, SVC $P=0.017$, left amygdala: *x*, *y*, $z=-24$, −6, −18, $t_{(23)}=3.78$, $Z=3.30$, SVC $P=0.018$] during the evaluation of the trustworthiness of Superiors than Inferiors in the decision phase (Fig. 3B). We then further explored whether this neural coupling was associated with participants' behavioral performance related to the perception of the trustworthiness of Superiors and Inferiors. This functional connectivity between right mPFC and left amygdala showed a significant correlation with pretest ratings of trustworthiness for Inferiors ($r=-0.46$, $P=0.024$).

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

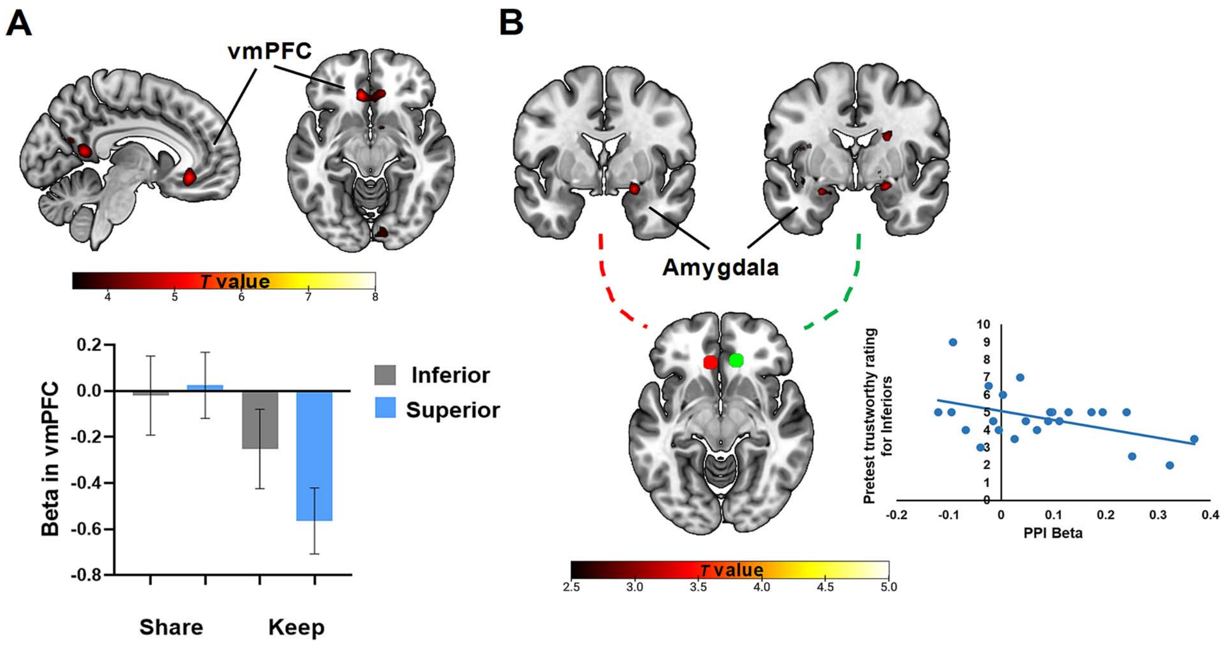

Fig. 3. Neural activity in the decision phase. A) Neural activation in the vmPFC was significantly higher when individuals chose to share with Superiors than chose to keep with Superiors (FWEc $P < 0.05$). B) Functional connectivity between the vmPFC and amygdala was stronger for the decision phase when interacting with Superiors than with Inferiors (SVC $P < 0.05$). This functional connectivity between right mPFC and left amygdala showed a significant correlation with pretest ratings of trustworthiness for inferiors ($r = -0.46$, $P = 0.024$). vmPFC, ventromedial prefrontal cortex;$*$ $P < 0.05$; error bars refer to SEM.

Feedback phase

For the feedback phase, we first sought to explore the clue that links social status and different feedback. The right amygdala revealed stronger activation when participants received feedback of reciprocity from Superiors than from Inferiors [x, y, z = 26, −6, −14, $t_{(23)} = 6.26$, $Z = 4.70$, SVC $P < 0.001$, Fig. 4A].

In the feedback phase, participants can update their beliefs of partners' trustworthiness and thus improve their decisions in the subsequent trials. Based on the reinforcement learning model (SV&LR Model), the updating is quantified by the PE. Using this model-derived trial-by-trial PE value as a parametric regressor, we identified the contribution of vmPFC [x, y, z = 6, 34, −14, $t_{(23)} = 7.19$, $Z = 5.16$, FWEc $P < 0.001$] and ventral striatum (VS) [x, y, z = −10, 8, −6, $t_{(23)} = 7.44$, $Z = 5.26$, FWEc $P < 0.001$] to encoding the updating of trustworthiness (Fig. 4B; Table 1). Next, we set out to characterize how social status influences the process by which participants place trust in their trustees during the trustworthiness updating. Thus, we contrasted the neural activation that correlated with the PE value in different partner contexts and found that the activation difference in vmPFC [x, y, z = −8, 44, −16, $t_{(22)} = 4.70$, $Z = 3.87$, SVC $P = 0.014$] and VS [VS, x, y, z = −14, 4, −12, $t_{(22)} = 5.53$, $Z = 4.33$, SVC $P = 0.001$] was significant. This result was observed when correlating the contrast of Superior minus Inferior with the model-based social value difference between participants (Fig. 4C). Then, we more closely examined this asymmetric modulation of trustworthiness updating by superior bias in a post hoc ROI analysis. In this analysis, one ROI was identified in vmPFC (with a 6 mm sphere centered on the peak coordinate x, y, z = −8, 44, −16), and another was identified in VS (with a 6 mm sphere centered on the peak coordinate x, y, z = −14, 4, −12). The PE-related BOLD activity was extracted from these cluster identifications. Results for the correlation between social value differences and parameter estimates of BOLD activation that parametrically vary with PE in the contrast of Superior minus Inferior were significant (vmPFC: $r = -\ 0.638$, $P < 0.001$; VS: $r = -\ 0.714$, $P < 0.001$). This finding indicates that the higher the social value of the Superior for participants, the lower the PE coding when they interacted with the Superior compared with the Inferior. In other words, prior-based superior bias in trust diminished the reliance on information updating in the neural circuitry of trial-and-error learning. The results of the computational modeling agreed with this finding, that superior bias in trust is negatively correlated with the difference in learning rates of the trustworthiness of Superior and Inferior partners.

Discussion

By combining computational modeling and fMRI, the current study determines the neurocomputational mechanisms that underlie how a trustee's social status influences the trust participants place in them. We implemented a multishot binary version of the trust game, in which trustors were required to make binary forced-choice decisions. Specifically, after receiving an endowment from the experimenter, the trustor had to decide whether to keep the entire amount (i.e., a distrust decision) or share it entirely (i.e., a trust decision) with trustees of either high or low social status. This version of the trust game categorizes the participants' choices as either trust or distrust on a qualitative basis, and the average percentage of trials in which trustors chose to share across experimental conditions serves as a measure of trust. The behavioral results revealed that participants trusted Superiors more than Inferiors when trustees showed similar trustworthiness. At the brain system level, social status influenced trust-related cognitive processing through two mechanisms: (i) the activity of the amygdala and vmPFC, and their functional connectivity, which represented the influence of social status on prior-based processing, and (ii) the dependency on the prior of social status, which diminished

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

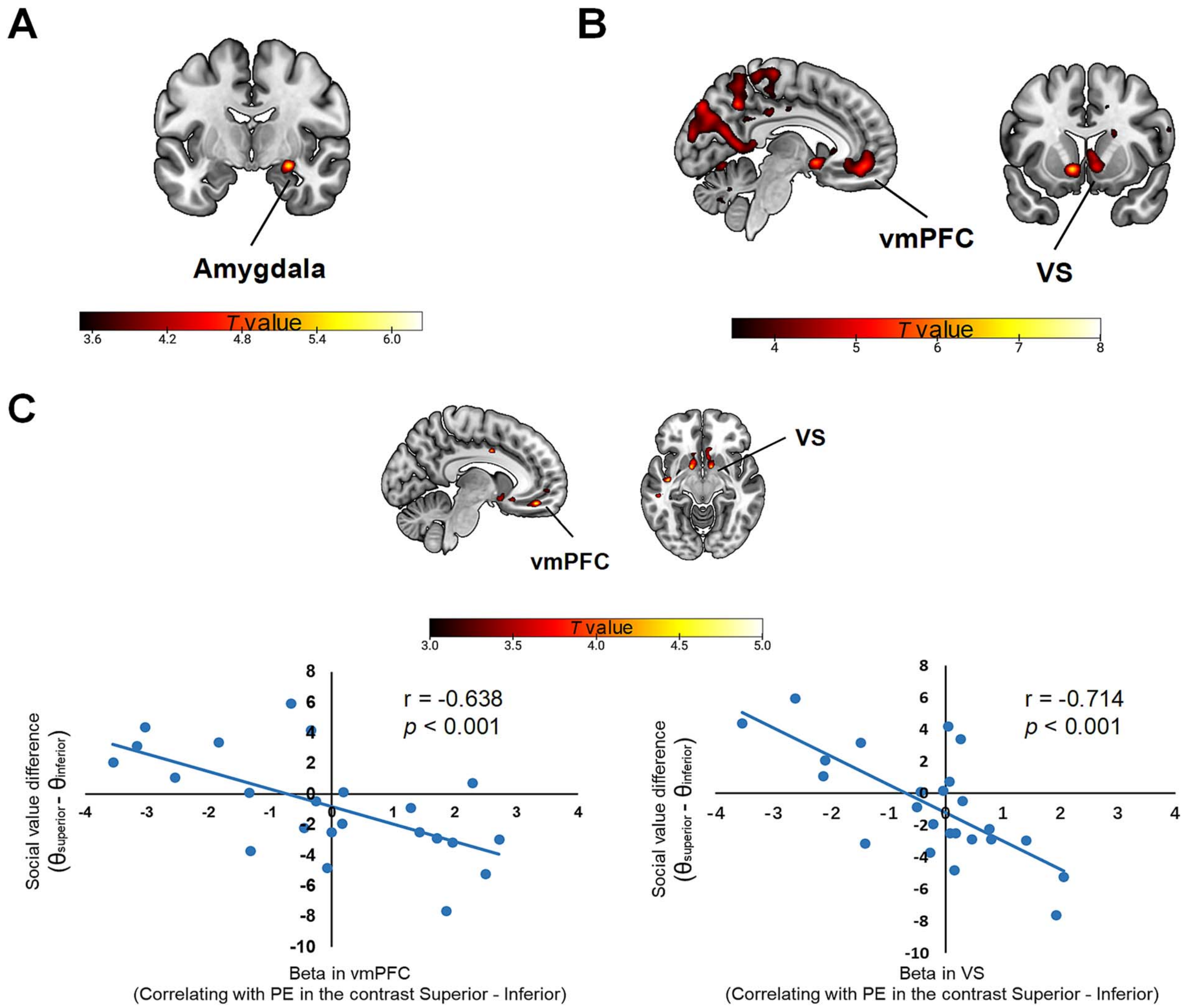

Fig. 4. Neural activity in the feedback phase. A) Neural activation in the right amygdala was significantly higher when participants received the feedback of a superior who was reciprocating than an inferior (SVC $P < 0.05$). B) Activity in vmPFC and VS shows a significant correlation with model-based trial-by-trial prediction error (FWEc $P < 0.05$). C) The subject-specific social value difference significantly modulated BOLD responses that correlated with prediction error in the contrast of Superior–Inferior in vmPFC and VS (SVC $P < 0.05$). The scatter plots illustrate the correlation between social value differences and parameter estimates of BOLD activation that parametrically vary with prediction errors in the contrast of Superior—Inferior, based on the post hoc ROI analyses of vmPFC and VS. vmPFC, ventromedial prefrontal cortex; VS, ventral striatum. PE, prediction error.

reliance on information updating based on PE engaging the vmPFC and VS, in the feedback-based dynamic processing.

Superior bias in trust-related behavioral representation

In our experiment, two kinds of information contributed to trust decisions: prior information about social status and trial-by-trial feedback about reciprocity. Here, we discuss how these two kinds of information affect trust decisions and the way they integrate. Both the trustworthy ratings and trust behaviors indicated that Superiors were considered to be more trustworthy than Inferiors, which is pervasive and relatively stable. In the first block of the trust game, when partners' reciprocity rates were unknown, participants had a stronger disposition to share with Superiors than Inferiors. However, over the course of consecutive trials, participants increasingly made decisions relying on trustees' reciprocity rates. Nevertheless, social status still had a significant influence, such that participants' share rates diverged when facing partners with the same reciprocity rate but different social statuses. This result is consistent with the finding of our previous study of a series of behavioral experiments, in which individuals with high status were more likely to gain the trust of others (Li et al. 2023).

Furthermore, social status was associated with trust and information updates. Feedback-based trial-and-error reinforcement learning is the main updating mechanism engaged in the repeated trust game, which has been delineated by an extensive body of work (Delgado et al. 2005; Chang et al. 2010; Fareri et al. 2012, 2015; Fouragnan et al. 2013; Konovalov et al. 2018). To explore the internal learning mechanism, we constructed different reinforcement learning models. The core updating mechanism, ie PE, contributes to quantifying the difference between the expected value and actual outcome. The learning rate is an important subject-specific parameter that quantifies updating by scaling PE (Ligneul et al. 2016, 2017; Konovalov et al. 2018; Lockwood and Klein-Flügge 2020; Zhang et al. 2020; Janet et al. 2022). We observed a negative correlation between the learning rate difference and the social value difference in the SV&LR Model, which revealed and quantified learning rates and social value for different social statuses. That is, the greater the difference in social value between high and low status in the participants' belief, the lower their

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

learning rates for the high-status partners compared with the low-status partners. A relatively lower learning rate for Superiors compared with Inferiors implies a weakened updating for Superiors among those who held the belief that there is a close relationship between high social status and high social value. Furthermore, we found a significant positive correlation between the difference in social value between high and low status and bias in favor of trusting Superiors. These results link two prejudices: that superior bias in trust may be driven by the high social value of status and that this subjective value influences information updating to stabilize the asymmetry of trust in the social status context. This finding is consistent with the second hypothesis we proposed concerning the impact of social status on trust learning. However, due to the absence of a discernible relationship between the learning rates of the outcomes of different valences and the social status (LG Model and SV&LG Model in the Supplementary material), we were unable to obtain results that could verify the first hypothesis.

Prior-based preference of social status affects trust-related processing

Social hierarchies are valuable social information that have a broad impact on social interactions (Zink et al. 2008; Fernald 2014; Santamaría-García et al. 2014; Hu et al. 2016; Qu et al. 2017). To explore the neural representation of social status in the context of trust reciprocity, we compared neural signals under different conditions of status and identified the important roles of vmPFC and amygdala in prior-based evaluation. Their roles may be related to the transmission and integration of social and value signals. We first explored the effect of social status on the decision phase by comparing the decision to share and keep. Our finding demonstrated that activity in vmPFC selectively modulates the decisions regarding superior but not inferior trustees. Specifically, only when playing with superior trustees, vmPFC showed stronger activation in decisions when they chose to trust rather than distrust. This relative activation was not observed in decisions involving trustees of inferior status. As a key structure of the valuation brain circuit, vmPFC is closely associated with encoding the expected value of stimuli; this guides decisions by incorporating complex and qualitatively different reward alternatives into a common currency of subjective value (Hare et al. 2008, 2009; Rangel et al. 2008; Sescousse et al. 2010; Louie and Glimcher 2012; Delgado et al. 2016). Based on previous research, the vmPFC may be sensitive to the higher social value associated with higher status.

Due to vmPFC's interconnectivity with the brain regions that support social and reward processing, including the amygdala and VS (Hare et al. 2009; Haber and Knutson 2010), the vmPFC functions as a neural integrator, integrating social signals via connectivity with several brain regions. Based on its structural features, we performed a PPI analysis with vmPFC as the seed region. The results showed that the functional connectivity between vmPFC and the amygdala varied according to the partners' social status. The amygdala was more strongly coupled to the vmPFC when participants interacted with Superiors than with Inferiors, and the strength of this functional connectivity was further correlated with the pretest ratings of trustworthiness for Inferiors. In other words, the participant's belief in social status may influence the cognitive neural mechanisms when interacting with partners of different ranks. If an individual is more capable of overcoming their belief that Inferiors are not trustworthy, in a trust game, there will be less disparity in the functional connectivity strength between mPFC and amygdala when interacting with partners of different ranks. In addition, the amygdala activity showed a stronger response when participants received reciprocation from Superiors than Inferiors in the feedback phase. Previous studies have revealed that neural activity in the amygdala can represent the value of others according to their rank in the social hierarchy. Thus, this signal could potentially be used to guide the selection of advantageous coalition partners (Kumaran et al. 2012; 2016). In trust-related processing, the amygdala evaluates the incoming social input and modulates behaviors by enhancing trust behaviors for positive evaluations or increasing distrust behaviors for negative evaluations (Koscik and Tranel 2011; Krueger and Meyer-Lindenberg 2019). People with damaged basolateral amygdala have difficulties in learning whom to trust in the trust game (Rosenberger et al. 2019). Associating the role of the amygdala in trust-related and social status-related processing, the amygdala may serve as a social stimulus evaluator that assesses the value of partners based on their social status and works together with the vmPFC to make the final trust-related decisions.

Feedback-based dynamic learning guides trust behaviors

In the repeated trust game, participants had the chance to update their beliefs with respect to others' trustworthiness and optimize their decisions. For the basic update mechanism, we found that neural activation of VS and vmPFC showed a significant correlation with PE, which replicated previous findings (Rangel et al. 2008; Olsson et al. 2020).

The subsequent primary focus of the current study centered on examining how social status, as a high-level social concept, would interact with the processing of feedback-based, trial-by-trial trustworthiness learning. Our results demonstrated that prior beliefs can reduce the impact of feedback-based updating. The subject-specific social value difference significantly modulated neural responses that correlated with PE in vmPFC and VS. Specifically, among participants, the higher the social value of Superiors compared to Inferiors, the lower the PE-related activity in the vmPFC and VS for Superiors compared with Inferiors. This finding further substantiates the second hypothesis. Our findings cohere with previous studies that show that knowledge about an interaction partner's prior knowledge can hinder or bias the updating of expectations, after interactions via the top–down modulation of reward circuitry (Delgado et al. 2005; Fareri et al. 2012; Fareri and Delgado 2014). One study provided a social prior by presenting visual cues and found that when no prior information was available, striatal activation patterns correlated with behaviorally estimated reinforcement learning measures. However, this correlation was disrupted when reputational priors of trustees were provided (Fouragnan et al. 2013). The violations of direct learning were mediated by prior-enhanced connectivity between the caudate nucleus and ventrolateral prefrontal cortex (Fouragnan et al. 2013). All of these findings suggest that prior information affects not only static cognitive processing but also dynamic and immediate information updating.

Limitations and future directions

Notwithstanding our insightful findings, there are several limitations that warrant discussion. Based on previous research findings, our correction of ROI, guided by a hypothesis-driven approach, adopts a relatively lenient methodology. The functions of these brain regions need to be continuously explored in future studies. Furthermore, in this study, we concentrated on investigating the impact of trustees' social status on their characteristics to gain trust from trustors, without considering

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

the interactive effects of social status between these two roles in trust-related interactions. This issue can be systematically examined in future research.

Conclusion

To conclude, we observed that participants trusted Superiors more than Inferiors when trustees showed similar trustworthiness, an effect likely due to Superiors holding an additional social value independent of trust profit, which thus resulted in superior bias (Li et al. 2023). Such social status influenced trust-related cognitive processing through two mechanisms at the time of choice and feedback. In particular, the social value bias modulated the updating mechanism at the time of feedback.

Acknowledgments

The authors wish to thank Dr. Sicong Liu for his invaluable advice regarding data analysis.

Author contributions

Siying Li (Conceptualization, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review & editing, Funding acquisition), Jean-Claude Dreher (Methodology, Writing—review & editing, Supervision), Edmund Derrington (Writing—review & editing), Haoke Li (Writing—review & editing), and Chen Qu (Conceptualization, Validation, Data curation, Supervision, Writing—review & editing, Funding acquisition).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This work was supported by the Ministry of Education of Humanities and Social Science Project (24YJC190014), the Program for National Natural Science Foundation of China (32171019), the MOE Project of Key Research Institute of Humanities and Social Sciences in Universities (22JJD190004), and Striving for the First-Class, Improving Weak Links and Highlighting Features (SIH) Key Discipline for Psychology in South China Normal University. This research also benefited from the financial support of IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005) and of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency. This work was also supported by grants from the Agence Nationale pour la Recherche to JCD (ANR-21-CE37–0032, ANR-24-EXEN-0001).

Conflict of interest statement. The authors declare no competing financial interest.

References

Adler NE, Epel ES, Castellazzo G, Ickovics JR. 2000. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy. *White Women Health Psychol*. 19:586–592. https://doi.org/10.1037/0278-6133.19.6.586.

Adolphs R, Tranel D, Damasio AR. 1998. The human amygdala in social judgment. *Nature*. 393:470–474. https://doi.org/10.1038/30982.

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Autom Control*. 19:716–723. https://doi.org/10.1109/TAC.1974.1100705.

Amaral DG. 2006. The amygdala, social behavior, and danger detection. *Ann N Y Acad Sci*. 1000:337–347. https://doi.org/10.1196/annals.1280.015.

Bartra O, McGuire JT, Kable JW. 2013. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*. 76: 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063.

Baumgartner T, Fischbacher U, Feierabend A, Lutz K, Fehr E. 2009. The neural circuitry of a broken promise. *Neuron*. 64:756–770. https://doi.org/10.1016/j.neuron.2009.11.017.

Bellucci G, Dreher JC. 2022. Neurocomputational signatures of learning to trust. In: *The neurobiology of trust* Krueger F (ed). Cambridge: Cambridge University Press, 10.1017/9781108770880.011.

Bellucci G, Chernyak SV, Goodyear K, Eickhoff SB, Krueger F. 2017. Neural signatures of trust in reciprocity: a coordinate-based meta-analysis: neural signatures of Trust in Reciprocity. *Hum Brain Mapp*. 38:1233–1248. https://doi.org/10.1002/hbm.23451.

Bellucci G, Molter F, Park SQ. 2019. Neural representations of honesty predict future trust behavior. *Nat Commun*. 10:5184. https://doi.org/10.1038/s41467-019-13261-8.

Berger J, Rosenholtz SJ, Zelditch M. 1980. Status organizing processes. *Annu Rev Sociol*. 6:479–508. https://doi.org/10.1146/annurev.so.06.080180.002403.

Blue PR, Hu J, Zhou X. 2018. Higher status honesty is worth more: the effect of social status on honesty evaluation. *Front Psychol*. 9:350. https://doi.org/10.3389/fpsyg.2018.00350.

Blue PR et al. 2020. Whose promises are worth more? How social status affects trust in promises. *Eur J Soc Psychol*. 50:189–206. https://doi.org/10.1002/ejsp.2596.

Bowles HR, Gelfand M. 2009. Status and the Evaluation of Workplace Deviance. *Psychol Sci*. 21:49–54. https://doi.org/10.1177/0956797609356509.

Chang LJ, Doll BB, van 't Wout M, Frank MJ, Sanfey AG. 2010. Seeing is believing: trustworthiness as a dynamic belief. *Cogn Psychol*. 61: 87–105. https://doi.org/10.1016/j.cogpsych.2010.03.001.

Cheng JT, Tracy JL, Foulsham T, Kingstone A, Henrich J. 2013. Two ways to the top: evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *J Psychol Soc Psychol*. 104:103–125. https://doi.org/10.1037/a0030398.

Delgado MR, Frank RH, Phelps EA. 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci*. 8:1611–1618. https://doi.org/10.1038/nn1575.

Delgado MR et al. 2016. Viewpoints: dialogues on the functional role of the ventromedial prefrontal cortex. *Nat Neurosci*. 19:1545–1552. https://doi.org/10.1038/nn.4438.

Delgado MR, Fareri DS, Chang LJ. 2023. Characterizing the mechanisms of social connection. *Neuron*. 111:3911–3925. https://doi.org/10.1016/j.neuron.2023.09.012.

Engell AD, Haxby JV, Todorov A. 2007. Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J Cogn Neurosci*. 19:1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508.

Fareri DS, Delgado MR. 2014. Social rewards and social networks in the human brain. *Neuroscientist*. 20:387. https://doi.org/10.1177/1073858414521869.

Fareri DS, Chang LJ, Delgado MR. 2012. Effects of direct social experience on trust decisions and neural reward circuitry. *Front Neurosci*. 6:148. https://doi.org/10.3389/fnins.2012.00148.

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

Fareri DS, Chang LJ, Delgado MR. 2015. Computational substrates of social value in interpersonal collaboration. *J Neurosci*. 35: 8170–8180. https://doi.org/10.1523/JNEUROSCI.4775-14.2015.

FeldmanHall O, Shenhav A. 2019. Resolving uncertainty in a social world. *Nat Hum Behav*. 3:426–435. https://doi.org/10.1038/s41562-019-0590-x.

Feldmanhall O et al. 2018. Stimulus generalization as a mechanism for learning to trust. *Proc Natl Acad Sci USA*. 115:E1690–E1697. https://doi.org/10.1073/pnas.1715227115.

Fernald RD. 2014. Communication about social status. *Curr Opin Neurobiol*. 28:1–4. https://doi.org/10.1016/j.conb.2014.04.004.

Foncelle A, Barat E, Dreher JC, Van der Henst JB. 2022. Rank reversal aversion and fairness in hierarchies. *Adapt Hum Behav Physiol*. 8: 520–537. https://doi.org/10.1007/s40750-022-00206-7.

Fouragnan E et al. 2013. Reputational priors magnify striatal responses to violations of trust. *J Neurosci*. 33:3602–3611. https://doi.org/10.1523/JNEUROSCI.3086-12.2013.

Gjoneska B, Liuzza MT, Porciello G, Caprara GV, Aglioti SM. 2019. Bound to the group and blinded by the leader: ideological leader–follower dynamics in a trust economic game. *R Soc Open Sci*. 6:182023. https://doi.org/10.1098/rsos.182023.

Haber SN, Knutson B. 2010. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*. 35:4–26. https://doi.org/10.1038/npp.2009.129.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A. 2008. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci*. 28:5623–5630. https://doi.org/10.1523/JNEUROSCI.1309-08.2008.

Hare TA, Camerer CF, Rangel A. 2009. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*. 324: 646–648. https://doi.org/10.1126/science.1168450.

Hare TA, Camerer CF, Knoepfle DT, O'Doherty JP, Rangel A. 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci*. 30:583–590. https://doi.org/10.1523/JNEUROSCI.4089-09.2010.

Hu J et al. 2016. Social status modulates the neural response to unfairness. *Soc Cogn Affect Neurosci*. 11:1–10. https://doi.org/10.1093/scan/nsv086.

Janet R et al. 2022. Regulation of social hierarchy learning by serotonin transporter availability. *Neuropsychopharmacology*. 47: 2205–2212. https://doi.org/10.1038/s41386-022-01378-2.

Joiner J, Piva M, Turrin C, Chang SWC. 2017. Social learning through prediction error in the brain. *Npj Sci Learn*. 2:8. https://doi.org/10.1038/s41539-017-0009-2.

Knutson B. 2005. Distributed neural representation of expected value. *J Neurosci*. 25:4806–4812. https://doi.org/10.1523/JNEUROSCI.0642-05.2005.

Konovalov A, Hu J, Ruff CC. 2018. Neurocomputational approaches to social behavior. *Curr Opin Psychol*. 24:41–47. https://doi.org/10.1016/j.copsyc.2018.04.009.

Koscik TR, Tranel D. 2011. The human amygdala is necessary for developing and expressing normal interpersonal trust. *Neuropsychologia*. 49:602–611. https://doi.org/10.1016/j.neuropsychologia.2010.09.023.

Kraus MW, Piff PK, Keltner D. 2011. Social class as culture: the convergence of resources and rank in the social realm. *Curr Dir Psychol Sci*. 20:246–250. https://doi.org/10.1177/0963721411414654.

Krueger F, Meyer-Lindenberg A. 2019. Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends Neurosci*. 42:92–101. https://doi.org/10.1016/j.tins.2018.10.004.

Krueger F et al. 2007. Neural correlates of trust. *Proc Natl Acad Sci*. 104:20084–20089. https://doi.org/10.1073/pnas.0710103104.

Kumaran D, Melo HL, Duzel E. 2012. The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron*. 76:653–666. https://doi.org/10.1016/j.neuron.2012.09.035.

Kumaran D, Banino A, Blundell C, Hassabis D, Dayan P. 2016. Computations underlying social hierarchy learning: distinct neural mechanisms for updating and representing self-relevant information. *Neuron*. 92:1135–1147. https://doi.org/10.1016/j.neuron.2016.10.052.

Li S, Krueger F, Camilleri JA, Eickhoff SB, Qu C. 2021. The neural signatures of social hierarchy-related learning and interaction: a coordinate- and connectivity-based meta-analysis. *NeuroImage*. 245:118731. https://doi.org/10.1016/j.neuroimage.2021.118731.

Li S, Huang G, Ma Z, Qu C. 2023. Superior bias in trust-related decisions. *Curr Psychol*. 42:24822–24836. https://doi.org/10.1007/s12144-022-03567-0.

Lieberman MD, Straccia MA, Meyer ML, Du M, Tan KM. 2019. Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): causal, multivariate, and reverse inference evidence. *Neurosci Biobehav Rev*. 99:311–328. https://doi.org/10.1016/j.neubiorev.2018.12.021.

Ligneul R, Obeso I, Ruff C, Dreher JC. 2016. Dynamical representation of dominance relationships in the human rostromedial prefrontal cortex. *Curr Biol*. 26:1–9. https://doi.org/10.1016/j.cub.2016.09.015.

Ligneul R, Girard R, Dreher JC. 2017. Social brains and divides: the interplay between social dominance orientation and the neural sensitivity to hierarchical ranks. *Sci Rep*. 7:45920. https://doi.org/10.1038/srep45920.

Liu X, Hairston J, Schrier M, Fan J. 2011. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*. 35:1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012.

Lockwood P, Klein-Flügge M. 2020. Computational modelling of social cognition and behaviour-a reinforcement learning primer. *Soc Cogn Affect Neurosci*. 16:761–771. https://doi.org/10.1093/scan/nsaa040.

Louie K, Glimcher PW. 2012. Efficient coding and the neural representation of value: Louie & Glimcher. *Ann N Y Acad Sci*. 1251: 13–32. https://doi.org/10.1111/j.1749-6632.2012.06496.x.

Ma DS, Correll J, Wittenbrink B. 2015. The Chicago face database: a free stimulus set of faces and norming data. *Behav Res Methods*. 47:1122–1135. https://doi.org/10.3758/s13428-014-0532-5.

Magee JC, Galinsky AD. 2008. 8 social hierarchy: the self-reinforcing nature of power and status. *Acad Manag Ann*. 2:351–398. https://doi.org/10.5465/19416520802211628.

Olsson A, Knapska E, Lindström B. 2020. The neural and computational systems of social learning. *Nat Rev Neurosci*. 21:197–212. https://doi.org/10.1038/s41583-020-0276-4.

Piff PK, Kraus MW, Côté S, Cheng BH, Keltner D. 2010. Having less, giving more: the influence of social class on prosocial behavior. *J Pers Soc Psychol*. 99:771–784. https://doi.org/10.1037/a0020092.

Polman E, Pettit NC, Wiesenfeld BM. 2013. Effects of wrongdoer status on moral licensing. *J Exp Soc Psychol*. 49:614–623. https://doi.org/10.1016/j.jesp.2013.03.012.

Qu C, Ligneul R, Van der Henst JB, Dreher JC. 2017. An integrative interdisciplinary perspective on social dominance hierarchies. *Trends Cogn Sci*. 21:893–908. https://doi.org/10.1016/j.tics.2017.08.004.

Rangel A, Camerer C, Montague PR. 2008. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*. 9:545–556. https://doi.org/10.1038/nrn2357.

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

Rosenberger LA et al. 2019. The human basolateral amygdala is indispensable for social experiential learning. *Curr Biol*. 29:3532–3537.e3. https://doi.org/10.1016/j.cub.2019.08.078.

Ruff CC, Fehr E. 2014. The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci*. 15:549–562. https://doi.org/10.1038/nrn3776.

Santamaría-García H, Pannunzi M, Ayneto A, Deco G, Sebastián–Gallés N. 2014. 'If you are good, I get better': the role of social hierarchy in perceptual decision-making. *Soc Cogn Affect Neurosci*. 9:1489–1497. https://doi.org/10.1093/scan/nst133.

Sescousse G, Redoute J, Dreher JC. 2010. The architecture of reward value coding in the human orbitofrontal cortex. *J Neurosci*. 30:13095–13104. https://doi.org/10.1523/JNEUROSCI.3501-10.2010.

Sladky R, Riva F, Rosenberger LA, Van Honk J, Lamm C. 2021. Basolateral and central amygdala orchestrate how we learn whom to trust. *Commun Biol*. 4:1329. https://doi.org/10.1038/s42003-021-02815-6.

Todorov A, Duchaine B. 2008. Reading trustworthiness in faces without recognizing faces. *Cogn Neuropsychol*. 25:395–410. https://doi.org/10.1080/02643290802044996.

Todorov A, Engell AD. 2008. The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc Cogn Affect Neurosci*. 3:303–312. https://doi.org/10.1093/scan/nsn033.

Tzieropoulos H. 2013. The trust game in neuroscience: a short review. *Soc Neurosci*. 8:407–416. https://doi.org/10.1080/17470919.2013.832375.

van Honk J, Eisenegger C, Terburg D, Stein DJ, Morgan B. 2013. Generous economic investments after basolateral amygdala damage. *Proc Natl Acad Sci*. 110:2506–2510. https://doi.org/10.1073/pnas.1217316110.

Yang Z, Zheng Y, Yang G, Li Q, Liu X. 2019. Neural signatures of cooperation enforcement and violation: a coordinate-based meta-analysis. *Soc Cogn Affect Neurosci*. 14:919–931. https://doi.org/10.1093/scan/nsz073.

Yu M, Saleem M, Gonzalez C. 2014. Developing trust: first impressions and experience. *J Econ Psychol*. 43:16–29. https://doi.org/10.1016/j.joep.2014.04.004.

Yu H, Siegel JZ, Clithero JA, Crockett MJ. 2021. How peer influence shapes value computation in moral decision-making. *Cognition*. 211:104641. https://doi.org/10.1016/j.cognition.2021.104641.

Zhang L, Lengersdorff L, Mikus N, Gläscher J, Lamm C. 2020. Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Soc Cogn Affect Neurosci*. 15:695–707. https://doi.org/10.1093/scan/nsaa089.

Zink CF et al. 2008. Know your place: neural processing of social hierarchy in humans. *Neuron*. 58:273–283. https://doi.org/10.1016/j.neuron.2008.01.025.

Downloaded from https://academic.oup.com/cercor/article/35/7/bhaf191/8209467 by INIST-CNRS INEE INSB user on 23 July 2025

© The Author(s) 2025. Published by Oxford University Press. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Cerebral Cortex, 2025, **35**, bhaf191
https://doi.org/10.1093/cercor/bhaf191
Original Article