OXFORD

Cerebral Cortex, 2022, 1–14
https://doi.org/10.1093/cercor/bhac365
Original Article

Learning whom to cooperate with: neurocomputational mechanisms for choosing cooperative partners

Tao Jin[1,2,3,4,†], Shen Zhang[1,2,3,†], Patricia Lockwood[5,6,7], Iris Vilares[4], Haiyan Wu[8], Chao Liu[1,2,3,*], Yina Ma[1,2,9,*]

[1]State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China,
[2]Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing 100875, China,
[3]Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, 100875, China,
[4]Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN, 55455, United States,
[5]Centre for Human Brain Health and Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, B15 2TT, United Kingdom,
[6]Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, OX3 9DU, United Kingdom,
[7]Department of Experimental Psychology, University of Oxford, Oxford, OX2 6GG, United Kingdom,
[8]Centre for Cognitive and Brain Sciences and Department of Psychology, University of Macau, Taipa, Macau SAR, 519000, China,
[9]Chinese Institute for Brain Research, Beijing, 102206, China

*Corresponding author: Chao Liu, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19 Xinjiekouwai Street, Beijing 100875, China. Email: liuchao@bnu.edu.cn; Yina Ma, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19 Xinjiekouwai Street, Beijing 100875, China. Email: yma@bnu.edu.cn

†Tao Jin and Shen Zhang contributed equally to this work.

Cooperation is fundamental for survival and a functioning society. With substantial individual variability in cooperativeness, we must learn whom to cooperate with, and often make these decisions on behalf of others. Understanding how people learn about the cooperativeness of others, and the neurocomputational mechanisms supporting this learning, is therefore essential. During functional magnetic resonance imaging scanning, participants completed a novel cooperation-partner-choice task where they learned to choose between cooperative and uncooperative partners through trial-and-error both for themselves and vicariously for another person. Interestingly, when choosing for themselves, participants made faster and more exploitative choices than when choosing for another person. Activity in the ventral striatum preferentially responded to prediction errors (PEs) during self-learning, whereas activity in the perigenual anterior cingulate cortex (ACC) signaled both personal and vicarious PEs. Multivariate pattern analyses showed distinct coding of personal and vicarious choice-making and outcome processing in the temporoparietal junction (TPJ), dorsal ACC, and striatum. Moreover, in right TPJ the activity pattern that differentiated self and other outcomes was associated with individual differences in exploitation tendency. We reveal neurocomputational mechanisms supporting cooperative learning and show that this learning is reflected in trial-by-trial univariate signals and multivariate patterns that can distinguish personal and vicarious choices.

Key words: computational modeling; cooperation-partner selection; multivariate pattern; prediction errors; vicarious learning.

Introduction

Cooperation—2 or more individuals working together or helping each other to achieve a common goal—is critical for the success of societies (Fehr and Fischbacher 2003). Cooperation exists between family members and genetically unrelated individuals, as well as between tribes, cities, and nations (Stallen and Sanfey 2013; De Dreu et al. 2020). Choosing the right partners is essential for the establishment and maintenance of cooperation (Noë and Hammerstein 1994), as there is substantial individual variability in the willingness of people to cooperate (Hula et al. 2018; Li et al. 2022). Moreover, successful partner selection has been associated with greater access to resources, more efficient problem solving, and a significant reduction in energy costs. However, in everyday life, not only do we have to select good partners for ourselves, but we also often choose cooperation partners on behalf of others, henceforth "vicarious partner choice." For example, a human resource manager recruiting a new employee for the team, and a professor hiring a research assistant for a post-doc. Despite the importance of understanding the mechanisms that drive personal and vicarious partner choice, the neurocomputational basis remains poorly understood.

How do we decide whether to cooperate with someone or not? An individual's physical appearance (e.g. facial and body features) provides a rapid clue to infer the cooperativeness of others (Stirrat and Perrett 2010). However, such snapshot-like perceptual cues are often inaccurate and make people fail to continuously update their belief of others' cooperativeness in light of new information and dynamic changes (Takahashi et al. 2006). A more reliable and accurate way to know the cooperativeness of a partner and to select them for oneself or others is to learn from the consequences of iterated interactions with the particular partner (McAuliffe et al. 2019). For example, we would prefer a cooperative partner who often performs to maximize mutual benefits over self-interest, as we could predict their cooperation with a high probability. In contrast, those who have betrayed us in previous encounters would not be selected again, as we would have a low expectation of cooperation. Thus, through positive (cooperative) and negative (uncooperative) outcomes of social interactions with

Received: April 27, 2022. **Revised:** August 15, 2022. **Accepted:** August 17, 2022

someone, we learn and update our expectation of cooperativeness, which guides our future decisions.

The framework of reinforcement learning (RL) theory provides a biologically plausible theoretical account for understanding how people update their expectations between actions and outcomes over multiple decisions (Daw et al. 2006). In this framework, the discrepancy between actual and expected outcomes (i.e. prediction errors, PEs) drives learning and, in the context of cooperation, could quantify how we learn about the cooperativeness of others. Indeed, recent studies have shown that people learn about traits of other agents, such as generosity, trustworthiness, and social dominance, in a manner consistent with the RL theory (Fareri et al. 2015; Hackel et al. 2015; Ligneul et al. 2016). When applying the RL framework to social learning, we can design experimental paradigms with carefully matched self and other conditions to directly compare them and test whether the computational and neural mechanisms overlap or are distinct (Lockwood et al. 2020). Previous work has shown that people use PEs to learn for others in a manner similar to how they learn for themselves (Lindström et al. 2018; Aquino et al. 2020). However, the extent to which different learning parameters are used can also differ when learning for oneself and for others. For example, individuals process self-relevant information more rapidly (Ma and Han 2010; Harris et al. 2018; Lockwood et al. 2018), are less willing to take risks when rewarding themselves (Beisswanger et al. 2003), and are less tolerant to temporal delays in receiving self-benefiting rewards (Albrecht et al. 2011). The current study aimed to reveal whether similar or distinct behavioral and neural processes are involved in choosing cooperative partners for the self and for others.

In terms of neural implementation, both human and nonhuman primate studies suggest that activity in the anterior cingulate cortex (ACC) is associated with decision-making regarding cooperation versus defection in dyadic interactions (Rilling et al. 2002; Haroush and Williams 2015). The ACC region, especially the more ventral portions of the ACC (perigenual ACC, pgACC), has been shown to encode PEs under both self- and other-referenced frameworks (Silvetti et al. 2014; Hill et al. 2016) and play an important role in representing dyadic similarity between self and other (Feng et al. 2018; Lau et al. 2020). We thus expected that the pgACC would play a crucial role in updating learning about other's cooperativeness and would be commonly involved in self and other processing.

We also expected different neural activity patterns engaged in the self and other conditions. The temporoparietal junction (TPJ), dorsolateral prefrontal cortex (dlPFC), precuneus (PCUN), and dorsomedial prefrontal cortex (dmPFC) have been linked to decision-making on behalf of others compared with oneself (Braams et al. 2014a, 2014b; Wu et al. 2020). We therefore predicted stronger activity in these regions when choosing cooperation partners vicariously for others. Finally, the finding that the striatum specifically responded personal (but not vicarious) PEs (Morelli et al. 2015; Sul et al. 2015) led us to predict stronger PE encoding in the striatum when receiving (either positive or negative) outcomes for the self than for others. In addition, extensive work suggests that individual differences in empathy are associated with variability in vicarious processing and prosocial behaviors (Singer et al. 2004; Lockwood et al. 2016). Thus, we further examined individual differences in learning to choose cooperative partners, especially how it is linked to trait empathy.

Here, we examined how individuals select cooperative partners for themselves and vicariously for others and included carefully matched self and vicarious experimental conditions. We designed a cooperation-partner-choice task (Fig. 1) in which participants played a modified version of a prisoner dilemma (PD) game where they were provided with the opportunity to choose a cooperation partner for themselves (hereafter the "self" condition, Fig. 1A) and vicariously for another gender-matched stranger (hereafter the "other" condition, Fig. 1B), rather than cooperate or defect as in a traditional PD (see Fig. 1C for payoffs). Although it is under the frame of PD, the current cooperation-partner-choice task is essentially different from PD. Specifically, participants were not provided with the opportunity to choose cooperation or defection, rather participants cooperated by default, and they could only choose whom to cooperate with. Unbeknownst to the participants, we designed 2 types of partners. One type of partner chose to cooperate 70% of the time, whereas the other one chose to cooperate 30% of the time. We would expect that participants would select the cooperative partner over the uncooperative partner by trial-and-error learning consistent with RL models. Critically, in half of the trials, participants made partner-choice decisions for themselves, and in the other half of the trials, participants made choices for another participant (a gender-matched stranger). Such a design allowed direct comparison between the self and other conditions within the same participant so as to identify the behavioral and neural similarities and differences between personal and vicarious learning about others' cooperativeness.

Materials and methods

Participants

Thirty-two right-handed healthy participants were recruited as paid volunteers in the current study. They participated in a cooperation-partner-choice task and made decisions in a Siemens Trio 3-T MRI scanner (Siemens, Erlangen, Germany). All participants had normal or corrected-to-normal vision, no history of psychiatric or neurological disorders or medication, and were not majoring in psychology or economics. Four participants were excluded from the analysis due to technical failure ($n = 1$) or excessive head movement (>3 mm; $n = 3$). Behavioral and neural data analysis was conducted on the remaining 28 participants (12 males, age range: 19–27 years, mean age = 22.25 years, and SD = 2.40). All participants gave written informed consent before the start of the experiment. The study protocol was approved by the Institutional Review Board of Beijing Normal University, Beijing, China.

Experimental task

We examined the behavioral and neural mechanisms underlying learning to choose cooperative partners for the self and for another agent. Each participant and a gender-matched confederate came to the experiment at the same time. Participants completed a cooperation-partner-choice task: choosing partners for a revised version of prisoner dilemma (rPD) game. Instead of choosing to cooperate or defect, participants cooperated by default in the rPD, and they chose 1 of the 2 gender-matched players to cooperate with, for themselves (learning for the self and the participants received the earned points: the self condition, Fig. 1A) or for the confederate (vicarious learning and the confederate received the earned points: the other condition, Fig. 1B). Participants were informed that 2 players of each presented pair would differ in cooperation propensity and that they needed to choose based on their own judgment. Unbeknownst to the participants, 1 player of each pair was more cooperative and would choose to cooperate with a probability of 70% (in the case of partner choosing to cooperate, both earned 200 points, Fig. 1C)

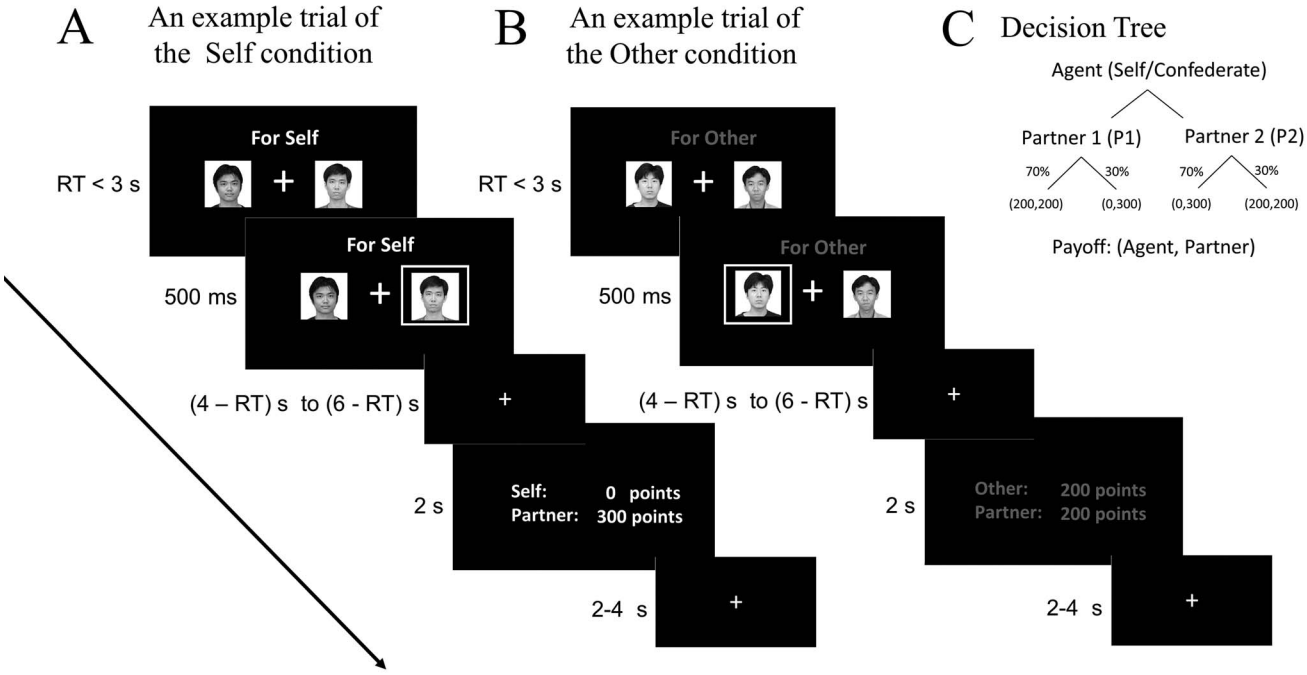

Fig. 1. Timeline and payoff matrix for the cooperation-partner-choice task. A, B) Timeline of the cooperation-partner-choice task. Participants were asked to choose 1 of 2 gender-matched partners for the self (the self condition, A) or another agent (the other condition, B). On each trial, participants were given 3 s to decide which 1 among 2 partners to cooperate with (the choice phase). After the decision was made, the chosen partner was highlighted for 500 ms followed by a fixation cross varying from 4 to 6 s minus RT. The chosen partner's decision and corresponding payoff were then presented for 2 s (the outcome phase), followed by an intertrial interval (varying from 2 to 4 s from a uniform distribution). If participants did not respond within 3 s during the choice phase, no decision was collected for that trial, and "no response" was presented during the outcome phase. A) An example trial of the self condition where participants chose a partner and received the outcome that the chosen partner had chosen to defect. B) An example trial of the other condition where participants chose a partner for another agent and received the outcome that the partner had chosen to cooperate. C) The payoffs of the revised PD game. If the chosen partner cooperates, the agent (either oneself or the other agent) and the chosen partner earn 200 points each. If the chosen partner defects, s/he earns 300 points, and the agent earns nothing. Points were converted into money at the end of the experiment. (RT: reaction time).

and the other player would choose to defect with a probability of 70% (in the case of partner choosing to defect, the partner earned 300 points whereas the participants/the confederate earned 0). To avoid potential effects of reciprocity, participants were told that the confederate would perform a different task irrelevant to their payoffs and that the money participants earned for the confederate would be given anonymously.

Participants encountered 4 pairs of players (with different face identities) and chose one of the players from each pair to cooperate with. Two of the pairs were used in the self condition (Fig. 1A), and 2 pairs in the other condition (Fig. 1B). There were 40 trials for each pair, resulting in a total of 160 trials (namely 80 trials in the self condition and 80 trials in the other condition). All pairs were randomly presented across 4 runs (40 trials/run), with no >2 trials of the same pair in a row. In each trial, photos of a particular pair of players were presented with the message "For Self"/"For Other," and participants were asked to choose 1 player as the partner for rPD within 3 s (the choice phase). The selected player was highlighted for 500 ms, followed by a fixation cross (varied between 4 and 6 s from a uniform distribution minus reaction time). Then, the payoff outcome of the trial was presented for 2 s (the outcome phase), followed by an intertrial interval that varied between 2 and 4 s from a uniform distribution. If participants did not respond within 3 s, this trial was terminated with a message "No response." The players in each pair were randomly assigned as cooperative or uncooperative players across participants and randomly presented on the left or right side of the screen across trials.

Procedure

Participants and a gender-matched confederate arrived at the functional magnetic resonance imaging (fMRI) center around the same time and completed the consent form and the interpersonal reactivity index (IRI; Davis 1983). Participants were then introduced to the cooperation-partner-choice task and performed 16 practice trials (8 trials in the self condition and 8 trials in the other condition). After being familiarized with the task, participants entered the scanner and performed the formal cooperation-partner-choice task. Participants completed pre- and postexperiment ratings of the cooperative propensity of all players on a 7-point Likert scale (1 = not cooperative at all, 7 = extremely cooperative). Participants' pre-experiment cooperativeness ratings were similar for all partners, but they rated cooperative partners as more cooperative than uncooperative partners after learning (Supplementary Fig. S1, see online supplementary material for a color version of this figure). Participants were paid for their participation (i.e. 20 United States dollar (USD) show-up fee) plus a bonus earned across 10 randomly selected trials in the self condition (100 points = 1 Chinese Yuan (CNY) ≈ 0.15 USD; i.e. ~2 USD).

Behavioral analysis

Behavioral data analysis

To test behavioral similarities and differences between the self and the other conditions, we compared the following indices between the self and other conditions: (i) the proportion of

choosing cooperative partners, (ii) learning criterion, i.e. the minimal-trial-number of choosing cooperative partners for 5 consecutive trials of the same pair, (iii) reaction time, and (iv) the proportion of choosing the same (cooperative and noncooperative) player in 2 consecutive trials of the same pair. We used a nonparametric test (i.e. Wilcoxon signed-rank test for the between-conditions comparison) if the distribution of the data violated the assumption of normality; otherwise, conventional paired sample *t*-tests were performed.

Computational modeling

RL models have been shown to capture the behavioral and computational mechanisms of reward learning and social learning (Daw et al. 2006; Fareri et al. 2015; Lockwood et al. 2016). In the current study, we employed RL models to understand how people learn to choose cooperative partners and whether the underlying computations were similar or distinct for the self and for another agent. Specifically, we built models based on the Rescorla–Wagner (R–W) value update rule to fit behavioral data and assessed the performance of these models.

First, we built a nonlearning model (M1) where we assumed that participants did not use partners' instant cooperation behaviors to guide their choices and thus set the learning rate α as 0 and inverse temperature β as a free parameter. Moreover, we fitted participants' choice data with 6 RL models. In these RL models, we tested whether the same or different computations underpinned participants' choices in the self and other conditions. We built and compared models that estimated shared parameters (i.e. M2, M4, and M6) or separate parameters (M3, M5, and M7) for the self and other conditions.

Model M2 is a basic RL model assuming same learning rate for all experiences. Specifically, we assumed that participants updated the expected value of an action of the chosen player *i* choosing to cooperate (EV_i). based on the difference (PE_i) between the expected and actual outcome R_i (coded as $R_i = 1$ or 0 if the chosen player chose to cooperate or defect, respectively) of the previous trial $t-1$. Specifically, participants updated the EV_i in trial *t* according to the Equations (1) and (2):

$$EV_i(t) = EV_i(t-1) + a * PE_i(t-1) \tag{1}$$

$$PE_i(t-1) = R_i(t-1) - EV_i(t-1) \tag{2}$$

where EV_i represents the expected value of an action that the chosen players choose to cooperate; the learning rate $\alpha \in [0,1]$ characterized the extent to which the EV_i is changed by PE_i ; PE_i represents the PE: The difference between the actual outcome (R_i) and the expected value of cooperation for the chosen player *i*. A high learning rate (α) indicated that the expected value of cooperation was volatile and largely influenced by the PE. We then employed the softmax function to transform the expected value into the probability of choosing a given player *i* in the Equation (3):

$$P_i(t) = \frac{\exp(\beta * EV_i(t))}{\exp(\beta * EV_i(t)) + \exp(\beta * EV_j(t))} \tag{3}$$

where the inverse temperature parameter β is a free parameter capturing the amount of exploration, i.e. the degree to which participants decides to choose a higher expected value choice versus exploring the other option. A low inverse temperature parameter β indicates that the participant has a similar likelihood of choosing either player irrespective of the expected value and that the choices are close to random. A high β suggests that the participant's choice is consistent and strongly driven by a higher expected value.

Model M3 is based on Model M2 except assuming different learning rates and inverse temperatures for the self and other conditions. The expected value is updated according to the Equation (4) for the self and the Equation (5) for the other agent:

$$EV_i(t) = EV_i(t-1) + \alpha_s * PE_i(t-1) \tag{4}$$

$$EV_i(t) = EV_i(t-1) + \alpha_o * PE_i(t-1) \tag{5}$$

where α_s and α_o are the learning rates for the self and the other agent, respectively. The expected values were transformed using the Equation (6) for the self and the Equation (7) for the other agent:

$$P_i(t) = \frac{\exp(\beta_s * EV_i(t))}{\exp(\beta_s * EV_i(t)) + \exp(\beta_s * EV_j(t))} \tag{6}$$

$$P_i(t) = \frac{\exp(\beta_o * EV_i(t))}{\exp(\beta_o * EV_i(t)) + \exp(\beta_o * EV_j(t))} \tag{7}$$

In Model M4, we included separate positive (cooperate) and negative (defect) learning rates. We assumed that participants learned about players' cooperative and defective behavior asymmetrically; hence, we used separate learning rates for the different outcomes as in the Equations (8) and (9):

$$EV_i(t) = EV_i(t-1) + \alpha^{[c]} * PE_i(t-1) \quad \text{if the chosen partner i cooperates in trial } (t-1) \tag{8}$$

$$EV_i(t) = EV_i(t-1) + \alpha^{[D]} * PE_i(t-1) \quad \text{if the chosen partner i defects in trial } (t-1) \tag{9}$$

Model M5 is based on Model M4 except assuming different learning rates and inverse temperatures for the self and other conditions, as in Equations (4–7).

Model M6 is based on Model 2. We further included a noise parameter (lapse ϵ [0,1]) to capture choice noisiness that is irrelevant to expected value differences (Equation (10)), allowing us to examine whether it is necessary to account for choice noisiness driven by factors independent of expected value differences (such as inattention):

$$P_i(t) = \left[\frac{\exp(\beta * EV_i(t))}{\exp(\beta * EV_i(t)) + \exp(\beta * EV_j(t))}\right] * (1 - \text{lapse}) + \frac{\text{lapse}}{2} \tag{10}$$

Model M7 is similar M6 except assuming different lapse, learning rates, and inverse temperatures for the self and other conditions, as in Equations (11) and (12):

$$P_i(t) = \left[\frac{\exp(\beta_s * EV_i(t))}{\exp(\beta_s * EV_i(t)) + \exp(\beta_s * EV_j(t))}\right] * (1 - \text{lapse}_s) + \frac{\text{lapse}_s}{2} \tag{11}$$

$$P_i(t) = \left[\frac{\exp(\beta_o * EV_i(t))}{\exp(\beta * EV_i(t)) + \exp(\beta_o * EV_j(t))}\right] * (1 - \text{lapse}_o) + \frac{\text{lapse}_o}{2} \tag{12}$$

Similar to previous work (Huys et al. 2011; Guitart-Masip et al. 2012; Mkrtchian et al. 2017), we employed a hierarchical Bayesian model fitting approach to estimate parameters. An expectation–maximization (EM) algorithm was used to find the maximum a

posteriori (MAP) estimate of each parameter for each participant (Huys et al. 2011). The EM algorithm started with the maximum likelihood estimation (MLE) for individual's parameters. In each iteration, a gaussian distribution (with mean and variance as free parameters) was estimated from individuals' parameters of the previous iteration. This gaussian distribution was then used as the prior for a MAP process to estimate individuals' parameters again. The iteration ceased when all parameters barely change. Compared with the widely used MLE method, this fitting method leverages individuals' parameters to estimate a group prior so as to regularize all individuals' parameters to the group mean and therefore less susceptible to extreme parameter values and guarantee a more accurate group estimate. Moreover, this fitting procedure has been thoroughly verified on the simulated data generated from the known decision process (Huys et al. 2011). To constrain parameters within meaningful ranges, exponential transforms were applied to the inverse temperature ($\geq$0), and the sigmoid transform was applied to the learning rate and lapse (varied between 0 and 1). These transformations indicate that the parameters are not normally distributed. Model fit was compared by the integrated Bayesian information criterion (iBIC). The iBIC was the integral of the likelihood function over the individual parameters. A lower iBIC indicates a better fit of the observed data.

To qualify the ability of the winning model to explain participants' behaviors, we ran a simulation analysis using the fitting parameters from the winning model. Specifically, for each subject, the set of fitting parameters was seen as a virtual subject and used to simulate choices of both the self and other conditions 100 times. The learning criterion and the stay probability of both conditions were computed for simulation and then were averaged across simulations to obtain a dataset of simulated behaviors. This simulated learning criterion and stay probability were then submitted to paired *t*-tests and were correlated with participants' actual behaviors to test whether the winning model successfully reproduced our model-free effects.

fMRI analysis

fMRI data acquisition and preprocessing

Whole-brain imaging data were acquired on a Siemens Trio 3-T MRI scanner with a 12-channel head coil at Beijing Normal University Brain Imaging Center, Beijing, China. Functional images were obtained using a $T2^*$-sensitive gradient-echo-planar imaging (EPI) sequence (33 slices; slice thickness = 3.5 mm, gap between slices = 0.7 mm, time repetition, TR = 2,000 ms; time echo, TE = 30 ms; field of view (FOV) = 224 mm; flip angle = 90°; and voxel size = 3.5 × 3.5 × 3.5-mm^3 spatial resolution). Structural images were collected using T1-weighted magnetization prepared rapid acquisition with gradient-echo (MPRAGE) sequence (144 slices; slice thickness = 1.33 mm; gap between slices = 0.66 mm; TR/TE = 2,530 ms; TE = 3.39 ms, FOV = 224 mm; flip angle = 7°; and voxel size = 1.3 × 1.3 × 1.3 mm^3 spatial resolution). Neuroimaging data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The first 4 volumes from each run were discarded to account for T1 equilibrium effects. Images were corrected for slice acquisition timing within each volume and realigned to correct head motions. Next, the realigned images were coregistered to the individual gray matter image segmented from the corresponding T1-weighted image and then normalized to Montreal Neurological Institute (MNI) space (resampled to 3 × 3 × 3 mm^3). Finally, a Gaussian kernel of 6-mm full-width at half-maximum was applied to spatially smooth the images.

Univariate activation analysis

After preprocessing, we created a generalized linear model (GLM) of blood oxygen level-dependent responses to examine brain regions in which neural activity was associated with latent variables (i.e. trial-wise expected value, PE) derived from the computational model (i.e. the winning model M3). In this GLM, missed trials, where participants did not make a choice, were included as a regressor of no interest. All regressors were convolved with a canonical hemodynamic response function. Six head motion parameters were also modeled to capture potential movement-related artifacts. A high pass filter with a cutoff of 128 s was employed.

The GLM integrated parameter estimates (i.e. trial-wise expected value, PE) derived from the computational model (i.e. the winning model M3). For each participant, we estimated a GLM with the following regressors of interest: (R1) a stick function at the onset of the choice phase in the self condition; (R2) R1 modulated by the expected value of cooperation in the self condition; (R3) a stick function at the onset of the outcome phase in the self condition; and (R4) R3 modulated by PEs in the self condition. Regressors 5–8 were similar to R1—R4, except that they were created for the other condition. Group mean estimated parameters in each condition were used to regularize the individual estimates and avoid noisy fitting by following previous studies (Seymour et al. 2012; Eldar et al. 2016). First-level contrast images were separately entered into a second-level random analysis to identify brain areas encoding the expected value (R2 and R6) and tracking PEs (R4 and R8) for self and for others, respectively. To determine the common neural representation of PEs, we performed a conjunction analysis (self condition ∩ other condition) within the flexible factorial framework (Nichols et al. 2005). To examine the distinct representations of expected value (PEs) between the self and other conditions, we conducted contrasts between R2 and R6 (between R4 vs. R8 for PE).

Statistical inference in both GLMs was performed at a standard threshold of $P < 0.05$, family-wise error (FWE) cluster-level corrected at the whole-brain level with a cluster-forming threshold of voxel-wise $P < 0.001$. We also performed $P < 0.05$ FWE small volume corrected (SVC) with an initial voxel-wise threshold of $P < 0.001$ for regions where we had a strong a prior hypothesis of encoding PEs (i.e. pgACC and striatum, Supplementary Fig. S2, see online supplementary material for a color version of this figure). We defined the pgACC based on a previous study examining how people learn about social information (Lau et al. 2020). The striatum was defined from term-based meta-analysis of "prediction error" in Neurosynth (Yarkoni et al. 2011).

Multi-voxel pattern analysis

Next, multi-voxel pattern analysis (MVPA) was employed to uncover finer-grained and spatially distributed neural activity underlying personal and vicarious learning to choose cooperative partners. Specifically, we examined the pattern of neural activity that could distinguish self-regarding from other-regarding choices (MVPA at the onset of the choice-making phase, MVPA 1) and outcomes (MVPA at the onset of outcome-presenting phase, MVPA 2). MVPA was implemented using slice-timing corrected and spatially realigned (but non-normalized and unsmoothed) images in the Decoding Toolbox (Hebart et al. 2015).

To assess activity patterns that discriminated self- and other-regarding choices (outcomes), we first estimated a GLM for each participant: The choice (outcome) on each trial of the self and other conditions was modeled as a single regressor at the onset of

choice-making (outcome-presentation) phase for MVPA 1 (MVPA 2). The GLM also included 2 regressors for outcomes (choices) in self and other conditions separately, 6 head-motion parameters and a regressor of missed trial as effects of no interest. Consistent with previous studies (Zhu et al. 2019), the estimated beta images of all choices were entered into a support vector machine (SVM) classifier with the leave-one-run-out cross-validation method. We used the default SVM classifier in the Decoding Toolbox, i.e. a SVM with linear kernel and a L2 regularization with the penalty parameter $C = 1$. A whole-brain searchlight decoding analysis was then performed using a sphere with a radius of 4 voxels. Beta values of each voxel were normalized across all pattern vectors by removing the mean and dividing the standard deviation. Then the beta values of N voxels in a given sphere were used to construct an N-dimensional pattern vector. The pattern vectors from 3 of 4 runs were used to train the SVM to discriminate between self- and other-regarding choices (outcomes for the second MVPA) and applied to the test run to obtain the classification accuracy of the test run. This process was iterated for the other 3 runs to calculate the mean cross-validated classification accuracy for each voxel, yielding a 3D map of classification accuracy. The individual accuracy maps were then spatially normalized and smoothed using the same parameters as those in the univariate activation analysis. At the group-level, these maps were then entered into a 1-sample t-test against chance level (50%). Multiple comparisons across the whole-brain were performed the same as the univariate activation analysis did, except the initial cluster-forming threshold of $P < 0.0001$ in MVPA 1.

Finally, we investigated the relationship between the neural activity pattern and behavioral differences in personal and vicarious learning about cooperation. The results of behavioral modeling showed a significant difference in the inverse temperature between the self and other conditions. We, therefore, extracted classification accuracy from a 6-mm sphere centered at the reported peak coordinates and calculated the correlations between the differences in the inverse temperature and the averaged classification accuracy of survived clusters. In addition, to examine whether the individual level of empathy modulates the other-self difference of both behaviors and neural responses, we calculated the correlation between the empathic concern subscale and the difference in other- and self-regarding learning behavior and neural underpinnings. We also regressed the behavioral and neural differences between self and other conditions on each of the IRI questionnaire subscales.

Results

In the cooperation-partner-choice task (Fig. 1), each participant encountered 4 pairs of gender-matched players in a revised PD game (2 pairs for the self condition, Figs. 1A; and 2 pairs for the other condition, Fig. 1B). In this revised PD, participants cooperated by default and could choose 1 player from each pair as the PD partner. If the partner also chooses to cooperate, both the partner and the participant (or the other agent) earn 200 points; if the partner chooses to defect, then the partner earns 300 points and the participant (or the other agent) earns 0 (Fig. 1C). In an event-related design, the 4 pairs were randomly presented (40 trials for each pair). In each trial, participants were presented with 2 players: 1 player chose to cooperate with a probability of 70% (cooperative partner), and the other player chose to cooperate with a probability of 30% (uncooperative partner). We expected that, through trial-and-error, participants would learn which player was the cooperative player for each pair of players.

Participants can accurately learn for self and other but are faster to choose cooperative partners for self

All results use parametric tests when data are normally distributed and nonparametric tests when data violate assumptions of normality (i.e. the Wilcoxon Signed-Rank test; Lindström et al. 2018). We first examined the probability of choosing the cooperative over the uncooperative partner and revealed that participants were able to learn cooperativeness of the players significantly above chance for both the self and other conditions (self: $M \pm SE = 71.55\% \pm 3.04\%$, vs. 50%, $t(27) = 7.095$, $P < 0.001$, 95% confidence interval (CI): [15.32%, 27.79%]; and other: $M \pm SE = 68.22\% \pm 3.00\%$, vs. 50%, $t(27) = 6.079$, $P < 0.001$, 95% CI: [12.07%, 24.37%]; Fig. 2A), and to a similar extent when learning for the self and for the other agent ($t(27) = 0.981$, $P = 0.335$, 95% CI: [−3.63%, 10.29%]).

The results of the Bayesian analysis lent further support for the null hypothesis of a similar accuracy for choosing the cooperative partner when learning for the self and other (Bayes factor [H1:H2] = 4.324). However, we also revealed differences in the learning processes for the self and other conditions. We found that participants were faster to learn for the self. Specifically, we set a learning criterion of choosing the cooperative player over the uncooperative player in at least 5 consecutive trials. When comparing self and other on this criterion, participants required fewer trials to reach the criterion successfully for themselves (minimal-trial-number to reach learned criterion, self: 14.93 ± 1.62; other: 18.45 ± 1.82; $t(27) = -2.196$, $P = 0.037$, 95% CI: [−6.805, −0.230]; Fig. 2B). Moreover, we found that participants were more likely to stay with their choice of the last trial (t—1) in the current trial (t) when choosing for themselves (self: $M \pm SE = 74.64\% \pm 2.47\%$; and other: $M \pm SE = 67.40\% \pm 2.27\%$; $t(27) = 2.944$, $P = 0.007$, 95% CI: [2.19%, 12.28%]; Fig. 2C). This was especially true when the chosen player cooperated in the previous trial (self: $M \pm SE = 82.39\% \pm 2.72\%$; and other: $M \pm SE = 74.33\% \pm 2.84\%$; $t(27) = 3.089$, $P = 0.005$, 95% CI: [2.71%, 13.40%]; Supplementary Fig. S3A, see online supplementary material for a color version of this figure) but not when the chosen player defected in the previous trial (self: $M \pm SE = 62.03\% \pm 2.90\%$; and other: $M \pm SE = 56.51\% \pm 2.08\%$; $t(27) = 1.84$, $P = 0.077$, 95% CI: [−0.65%, 11.69%]; Supplementary Fig. S3B, see online supplementary material for a color version of this figure).

Interestingly, reaction time analyses showed that participants made decisions faster when choosing partners for the self than for another agent (self: 1.212 ± 0.038 s; and other: 1.259 ± 0.042 s; $t(27) = 2.817$, $P = 0.009$, 95% CI: [−0.083 s, −0.013 s]; Fig. 2D). Moreover, the differential reaction times when choosing for another agent and the self was associated with individuals' empathic concern, one of the IRI (Davis 1983) subscales. Individuals scoring higher on the empathic concern made decisions more slowly and deliberately when choosing for other people than for the self (Pearson's $r = 0.433$, $P = 0.021$, Fig. 2E). Multiple regression, including all IRI subscales, confirmed that the association between the self-other differential reaction time and empathy score was specific to the empathic concern subscale ($\beta = 0.011$, SEM = 0.004, $t = 2.564$, $P = 0.017$).

Next, we built models to fit participants' choice data and to reveal the computations underlying choosing cooperative partners, as well as to assess whether participants employed similar or distinct computations when choosing for self and other. We tested 7 models based on the Rescorla–Wagner (R–W) value update rule (Table 1): A baseline no-learning model (M1), basic R–W models (M2 and M3), R–W models with separate positive (cooperate) and

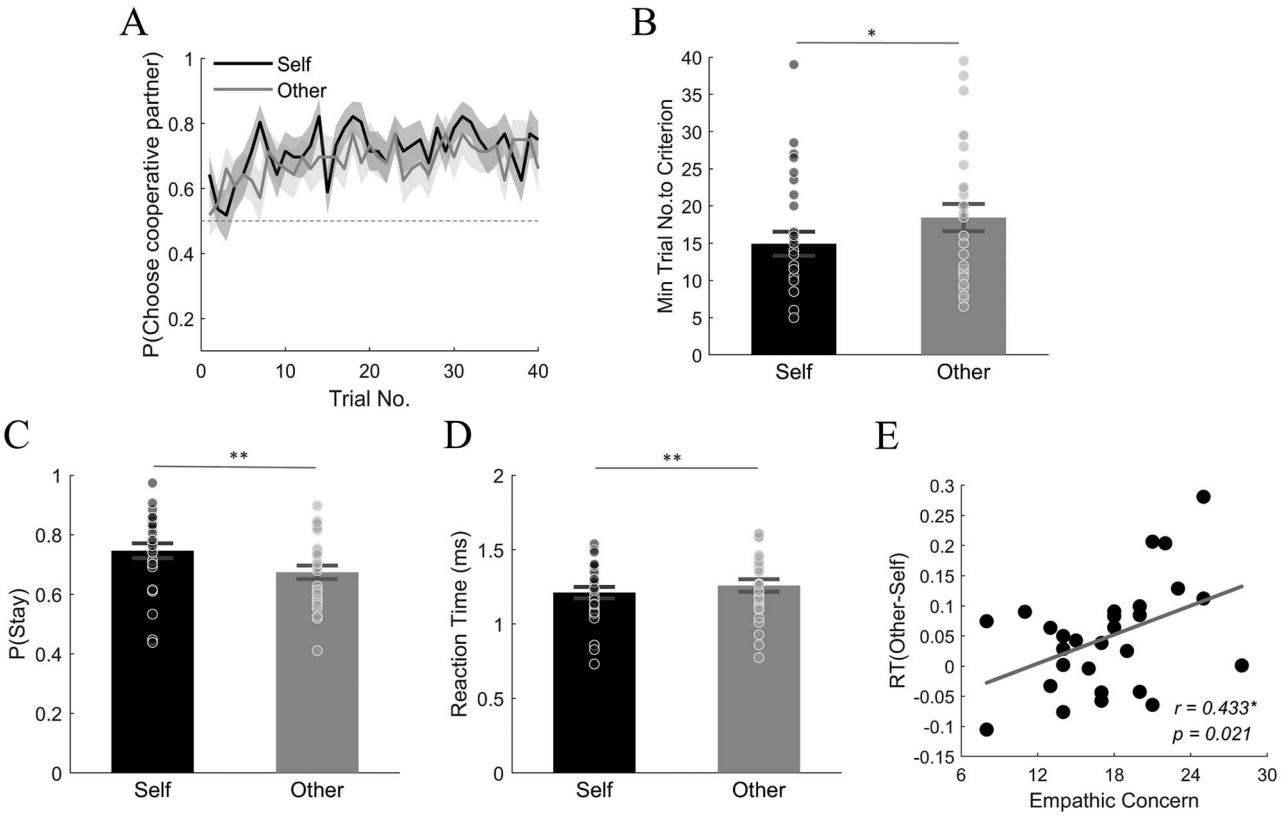

Fig. 2. Behavioral results. A) The group-level proportion of choosing cooperative partners changed over time in the self (blue) and the other (green) conditions. Dashed line shows the chance level of 50%. Lines and shaded areas show the mean ± SEM of choice proportion. B) The minimum number of trials that was required to reach the learning criterion was smaller in the self condition than in the other condition. C) A higher proportion of staying with the chosen partner in the previous trial, when choosing for self than for the other agent. D) Participant decided faster when choosing for the self than for the other agent. E) The positive association between trait empathic concern and the difference in reaction time when choosing for another agent and the self. (data are shown as the mean ± SEM with overlaid dot plots. * $P < 0.05$; ** $P < 0.01$).

negative (defect) learning rates (M4 and M5), and R–W models with choice noise being irrelevant to value difference (M6 and M7). Of particular interest, we compared the R–W models with single parameters shared by the self and other conditions (i.e. M1, M2, M4, and M6) and those with separate parameters split between the self and other trials (i.e. M3, M5, and M7).

Model comparison revealed that participants' choices of cooperative players were most parsimoniously explained by the basic R–W model with separate learning rates and inverse temperatures for the self and other conditions (i.e. M3; see Table 1 for the iBIC for all models tested). Here, the learning rate captures the extent to which participants update their choices based on recent feedback and the inverse temperature captures the extent to which participants made their choices based on the value difference between the 2 players. Model comparison results suggested that the same computational algorithm (i.e. the basic R–W learning strategy without featuring valance-sensitive learning rates) was employed to choose cooperative partners for the self, as well as for another agent. Moreover, individuals utilized different weighting of model parameters to support their choices when choosing for themselves and for another agent.

For sanity check, we used the set of individual fitting parameters derived from the winning model to simulate data mimicking participants' behaviors (detailed in Methods). The simulated data reproduced the same behavioral patterns of the learning criterion and the stay probability as participants showed: (i) smaller minimal-trial-number to reach learned criterion in the self than other conditions (self: 15.89 ± 0.60, other: 17.43 ± 0.70, $t(27) = -2.26$, $P = 0.032$, CI: [−2.95,-0.14], Fig. 3A) and (ii) larger stay probability for the self than other conditions (self: 68.67% ± 1.49%, other: 64.42% ± 1.57%, $t(27) = 2.84$, $P = 0.009$, CI: [1.18%, 7.33%], Fig. 3B). Moreover, the model simulated behaviors were highly consistent with participants' actual behaviors (minimal-trial-number, self: $r = 0.66$, $P < 0.001$, other $r = 0.75$, $P < 0.001$; Fig. 3C; stay probability, self: $r = 0.76$, $P < 0.001$, other: $r = 0.73$, $P < 0.001$; Fig. 3D). These results confirmed the capability of our wining model to explain participants' behaviors.

Next, we compared these parameters between self and other. We found that learning rates for the self and other conditions were not significantly different (self: $M \pm SE = 0.275 \pm 0.04$; other: $M \pm SE = 0.274 \pm 0.04$; $V = 212$, $P = 0.849$; Bayes factor [H1:H2] = 6.852; Fig. 3E). However, participants showed a significantly higher inverse temperature when choosing partners for the self than for another agent (self: $M \pm SE = 3.91 \pm 0.32$; other: $M \pm SE = 3.30 \pm 0.28$; $V = 295$, $P = 0.036$; Fig. 3F), suggesting that more exploitative decisions were made when choosing for the self. This was consistent with the model-free result of higher staying probability in the self condition.

Taken together, individuals updated their beliefs of other's cooperativeness using immediate evidence to a similar extent when choosing for oneself and another agent. However, individuals made decisions faster, relied more on the partners' cooperative tendency, and preferred to choose the previously cooperated partners when choosing partners for oneself (vs. others).

Table 1. Model summary.

Model no.	NP	Parameter								iBIC
M1	1	Inverse temperature								6,141.4
M2	2	Learning rate				Inverse temperature		—		4,694.1
M3	**4**	**LR (Self)**		**LR (Other)**		**IT (Self)**	**IT (Other)**	—		**4,667.5**
M4	3	$LR^{[C]}$		$LR^{[D]}$		Inverse Temperature		—		4,691.7
M5	6	$LR^{[C]}$ (Self)	$LR^{[C]}$ (Other)	$LR^{[D]}$ (Self)	$LR^{[D]}$ (Other)	IT (Self)	IT (Other)	—		4,669.7
M6	3	Learning Rate				Inverse Temperature		Lapse		4,705.3
M7	6	LR (Self)		LR (Other)		IT (Self)	IT (Other)	Lapse (Self)	Lapse (Other)	4,668.8

Note. Model no.: model number; NP: number of parameters; LR: learning rate; IT: inverse temperature; [C]: parameter for cooperate; [D]: parameter for defect; iBIC: integrated Bayesian information criterion.

fMRI results

Common and distinct PEs encoding for oneself and another agent

We first examined neural activity that encoded PEs during learning of cooperative partners, by correlating neural responses during the outcome phase with the trial-wise PE derived from the winning model. Previous studies have revealed that the pgACC is implicated in cooperative behaviors (Rilling et al. 2002; Haroush and Williams 2015) and encoding PEs (Silvetti et al. 2014; Hill et al. 2016). In addition, the striatum was reported in social RL (Chang and Sanfey 2009; Fareri et al. 2015) and reward-based learning (Lefebvre et al. 2017). We thus identified the pgACC and striatum as regions of interest (ROIs) and examined whether and how neural responses in the pgACC and striatum encoded self-regarding and other-regarding Pes via ROI analysis. The pgACC and striatum ROIs were adopted based on the results of previous studies examining learning about social coalition (Lau et al. 2020) and term-based meta-analysis of "prediction error" in Neurosynth (Yarkoni et al. 2011), respectively. Parameter estimates (β values) respectively associated with Pes in the self and other conditions were extracted and averaged over all voxels within each ROI.

We found that the activity in the pgACC activity co-varied with Pes in both the self and other conditions (self: $M \pm SE = 0.60 \pm 0.15$, $P < 0.001$, $t = 3.99$, 95% CI: [0.29, 0.90]; other: $M \pm SE = 0.57 \pm 0.16$, $P = 0.001$, $t = 3.55$, 95% CI: [0.24, 0.90]) and to a similar extent ($P = 0.89$, $t = 0.14$, 95% CI: [−0.35, 0.40]; Bayes factor [H1:H2] = 6.791, Fig. 4A). However, the bilateral striatum signaled Pes only in the self condition (left striatum: $M \pm SE = 0.68 \pm 0.16$, $P < 0.001$, $t = 4.18$, 95% CI: [0.35, 1.02]; right striatum: $M \pm SE = 0.73 \pm 0.17$, $P < 0.001$, $t = 4.35$, 95% CI: [0.39, 1.07], Fig. 4B and Supplementary Fig. S4A, see online supplementary material for a color version of this figure) but not in the other condition (left striatum: $M \pm SE = 0.33 \pm 0.20$, $P = 0.11$, $t = 1.65$, 95% CI: [0.35, 1.02]; self vs. other: $P = 0.072$, $t = 1.87$, 95% CI: [−0.03, 0.73]; right striatum: $M \pm SE = 0.27 \pm 0.19$, $P = 0.15$, $t = 1.47$, 95% CI: [0.03, 0.67]; self vs. other: $P = 0.014$, $t = 2.64$, 95% CI: [0.16, 0.75], Fig. 4B and Supplementary Fig. S4A, see online supplementary material for a color version of this figure). Interestingly, we found that the striatum activity that differently encoded the other and self Pes varied as a function of individual trait in the empathic concern subscale. Individuals scoring lower in empathic concern showed larger differences in striatum activity encoding other (vs. self) Pes (left striatum: Pearson's $r = 0.41$, $P = 0.030$; right striatum: Pearson's $r = 0.52$, $P = 0.005$, Fig. 4C and Supplementary Fig. S4B, see online supplementary material for a color version of this figure). The multiple regression, including all IRI subscales, also supported that this association was specific to the empathic concern subscale (left striatum: $\beta = 0.103$, SEM = 0.048, $t = 2.157$, $P = 0.042$; right striatum: $\beta = 0.115$, SEM = 0.042, $t = 2.763$, $P = 0.011$).

In addition, we conducted a whole-brain analysis and confirmed that the trial-by-trial Pes were encoded by activity in the pgACC, bilateral striatum, and precentral gyrus for the self condition (Fig. 4D) but only in the pgACC for the other condition ($P < 0.05$ FWE whole-brain corrected at the cluster-level after voxel-wise thresholding at $P < 0.001$; Fig. 4D, Supplementary Table S1). We next examined the shared and distinct encoding of PEs in the self and other conditions. A conjunction analysis (Nichols et al. 2005) confirmed the overlap in PE encoding in the pgACC for both the self and other conditions (FWE-SVC after voxel-wise thresholding at $P < 0.001$; Fig. 4E). The direct comparison of neural activity related to PE computations for self versus for others revealed stronger activity in the bilateral striatum encoding PE when learning for the self than the other agent (FWE-SVC after voxel-wise thresholding at $P < 0.001$; Fig. 4F). These findings suggested that when learning about other's cooperativeness, pgACC signaled personal and vicarious PEs, whereas the striatum specialized in coding PEs in regard to the self.

Neural patterns discriminating self versus vicarious choices and outcomes

We next employed multivariate pattern analyses (MVPA) to probe how personal and vicarious learning of cooperativeness was distinctly represented in spatially distributed neural response patterns in the brain. We aimed to reveal neural patterns distinguishing the self and other at both decision-making and outcome phases. We used a whole-brain searchlight to obtain a classification accuracy value per voxel. A whole-brain searchlight analysis at the onset of the choice phase, thresholding at a standard cluster-forming threshold of $P < 0.001$, resulted in a cluster of 147,090 voxels spanning a large portion of the brain (Supplementary Fig. S5, see online supplementary material for a color version of this figure). We therefore used a more stringent cluster-forming threshold of $P < 0.0001$ to obtain precise clusters that differentiated self and other choices. This analysis revealed that the neural activity pattern in the left dlPFC, the right TPJ, the occipital cortex comprising the left TPJ, the right middle temporal gyrus (MTG), the left superior/middle frontal gyrus (SFG/MFG), and the cerebellum differentiated choice-making for the self and other conditions ($P < 0.05$ FWE whole-brain corrected at the cluster-level after voxel-wise thresholding at $P < 0.0001$; Fig. 5A). The MVPA results showed that neural patterns in a range

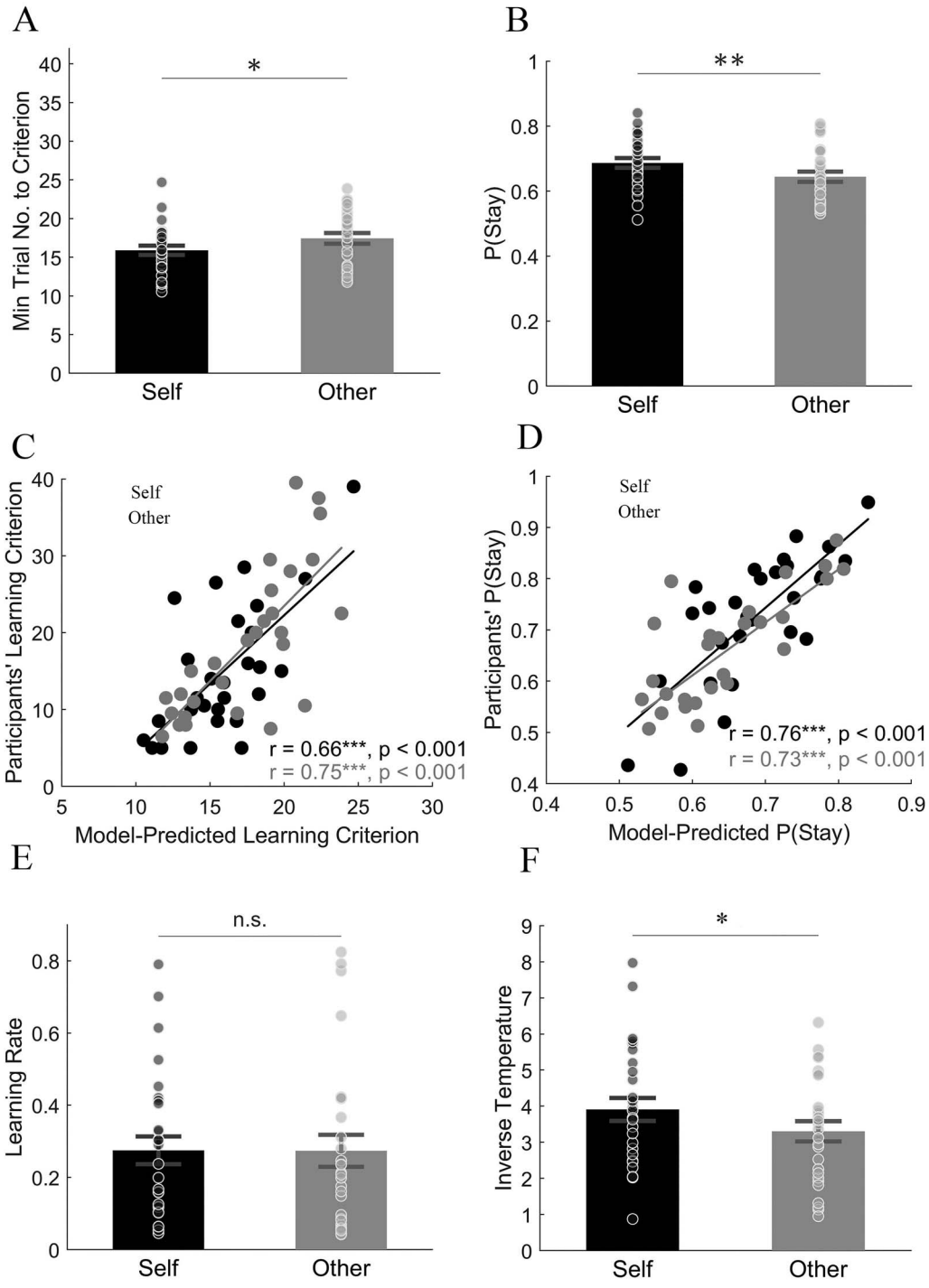

Fig 3. Computational modeling results. Simulated data using the winning model reproduced the effect of the minimum number of trials that was required to reach the learning criterion A), and the probability of staying with the chosen partner in the previous trial B). Moreover, the learning criterion C) and the probability of staying D) computed by simulated data were highly correlated with participants' actual behavior. E) Similar learning rates when learning cooperative partners for the self and for the other agent. F) The inverse temperature, characterizing the sensitivity toward value difference, was significantly larger in the self condition than in the other condition. (data are shown as the mean ± SEM with overlaid dot plots. N.s., nonsignificant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

of brain regions distinguished personal choices from vicarious choices during learning cooperativeness.

Next, we implemented another MVPA to reveal the neural activity patterns differentiating the processing of personal and vicarious outcomes. This analysis identified differential activity patterns in the right TPJ, the left striatum, the PCUN, the left superior temporal gyrus (STG), and cortical midline structures, including the posterior cingulate cortex (PCC) and the dorsal ACC ($P < 0.05$ FWE whole-brain-corrected at the cluster-level after voxel-wise thresholding at $P < 0.001$; Fig. 5B, Supplementary Table S2).

Finally, we tested for the association between our behavioral differences in choosing for the self and another agent (i.e. differences in the inverse temperature between self and other

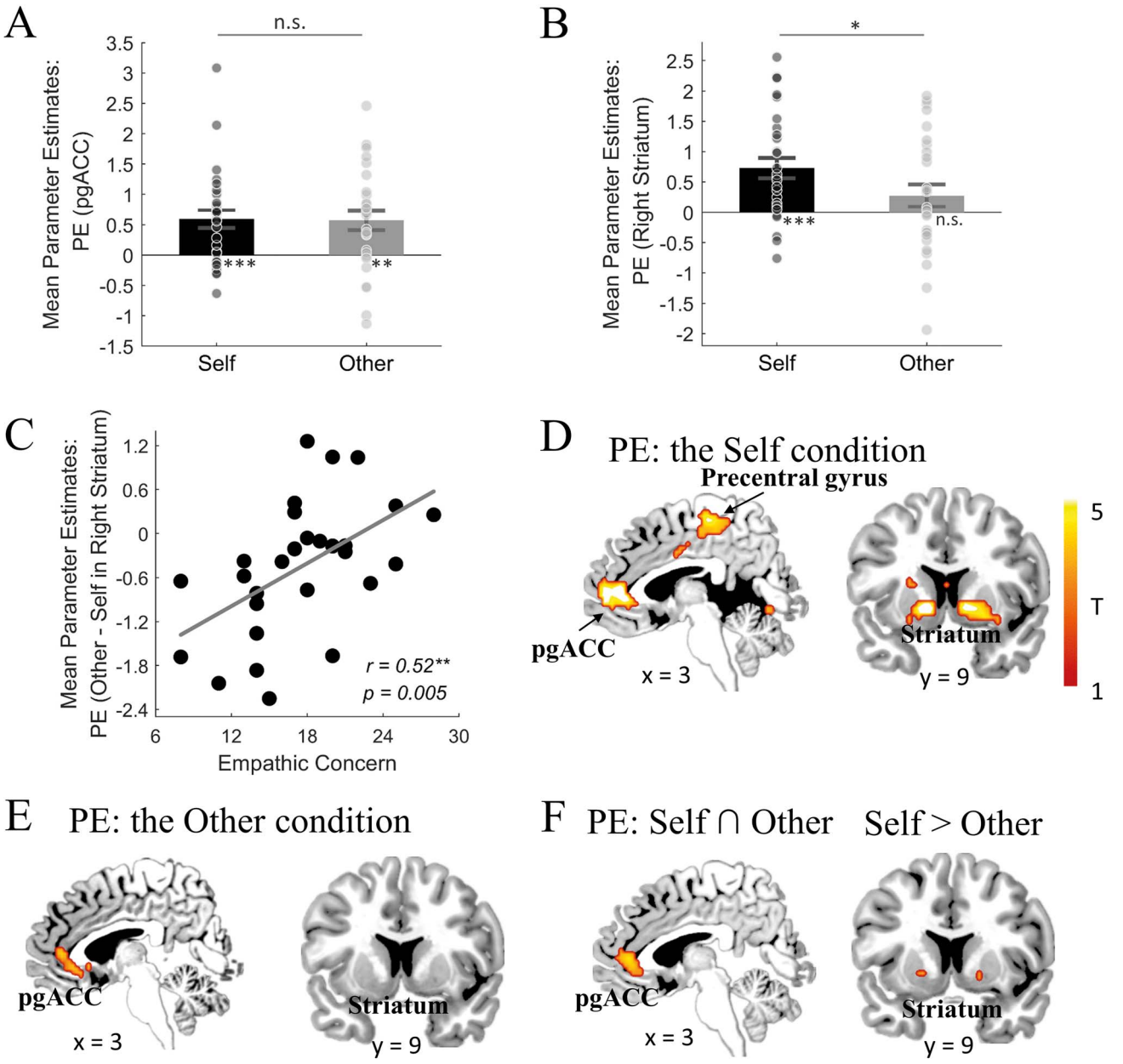

Fig 4. Common and distinct neural responses encoding PEs for self and the other agent. A, B) The ROI analysis revealed that the pgACC tracked PEs in both self and other conditions, but the right striatum signaled PE exclusively in the self condition. C) Scatterplot revealed that individual differences in empathic concern modulated the neural responses to other-regarding versus self-regarding PE in the right striatum. That is, the neural activity in the right striatum responded less to other-regarding PE relative to self-regarding PE for people lower in empathic concern. D) The pgACC (peak MNI coordinates [x, y, z]: [0, 42, 6], cluster size $k = 602$, $t = 6.68$), bilateral striatum ([−12, 9, −3], $k = 61$, $t = 6.30$) and precentral gyrus ([−6, −33, 54], $k = 641$, $t = 5.91$) signaled PEs in the self condition. E) The pgACC ([9, 30, 9], $k = 128$, $t = 4.69$) tracked PEs in the other condition. $P < 0.05$ FWE whole-brain corrected at the cluster-level after voxel-wise thresholding at $P < 0.001$ for panels D and E. F) Conjunction analysis revealed that pgACC activity tracked PEs regardless of the agent ([−3, 39, 0], $k = 104$, $P < 0.001$, $t = 4.65$). The contrast analysis revealed that the bilateral striatum tracked PEs to a greater degree in the self condition than in the other condition (left: [−18, 9, −3], $k = 2$, $P = 0.017$, $t = 3.75$; right: [18, 6, −6], $P = 0.015$, $k = 3$, $t = 3.60$). $P < 0.05$ FWE-SVC after voxel-wise thresholding at $P < 0.001$. Display threshold is $P < 0.001$ uncorrected. Error bars represent the standard error of the mean; n.s., nonsignificant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

conditions) and the neural responses to personal and vicarious cooperative learning. We found that larger differences in inverse temperature (but not learning rate) were associated with a higher classification accuracy of right TPJ in differentiating personal and vicarious outcome processing (Pearson's $r = 0.473$, $P = 0.011$, survives from False Discovery Rate (FDR)-correction for multiple comparisons, Fig. 5C), suggesting a consistent classification of learning for the self and the other agent at both behavioral and neural levels. Therefore, those participants with the largest self and other difference in learning were the same participants who distinguished self and other more at the time of the outcome in the right TPJ. It also should be cautious about our decoding results due to the relatively small sample size.

Discussion

Knowing whether someone is cooperative plays a crucial role in successful social life. The current study examined the neurocomputational mechanisms for how people learn to choose cooperative partners. We found that despite showing similar

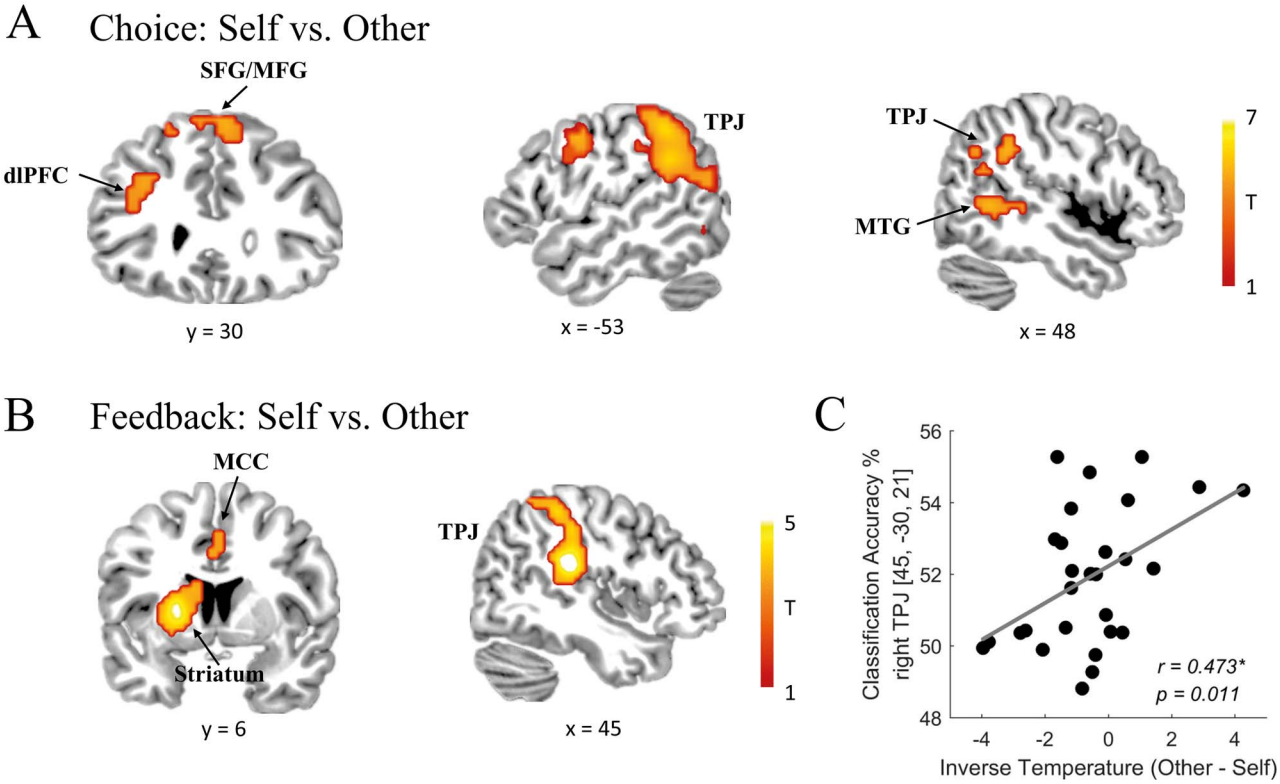

Fig 5. Differential neural activity patterns at the onset of choice and outcome phases. A) Using a whole-brain searchlight analysis, we found that the distributed activity patterns of the left dlPFC (peak MNI coordinates [x, y, z]: [−33, 30, 30], cluster size $k = 39$, $t = 4.93$), occipital cortex comprising the left TPJ ([0, −84, 6], $k = 4{,}300$, $t = 8.99$), right MTG extending into the right TPJ ([48, −51, 3], $k = 179$, $t = 5.24$) and SFG/MFG ([−18, 24, 54] $k = 326$, $t = 5.36$) could significantly distinguish between self- and other-regarding choices. The display threshold is $P < 0.0001$ uncorrected and significant activations are defined at $P < 0.05$ FWE whole-brain corrected at the cluster-level after voxel-wise thresholding at $P < 0.0001$. B) The distributed activity patterns of the left striatum ([−21, 6, 6], $k = 401$, $t = 5.14$), right TPJ ([45, −30, 21], $k = 543$, $t = 5.58$) and dorsal ACC ([0, 21, 45], $k = 526$, $t = 5.04$) could significantly distinguish between self- and other-regarding outcomes. The display threshold is $P < 0.001$ uncorrected and significant activations are defined at $P < 0.05$ FWE whole-brain corrected at the cluster-level after voxel-wise thresholding at $P < 0.001$. C) Bivariate association between the classification accuracy of self- and other-regarding outcomes in the right TPJ and the difference in inverse temperature between the other and self conditions (Pearson's $r = 0.473$, $P = 0.011$). Participants who behaved more exploitatively for other compared with the self had more distinct multivariate patterns in the right TPJ.

learning rates when choosing cooperative partners for oneself and vicariously for another person, people made slower and less exploitative choices for others than for themselves. This effect was modulated by individual levels of empathic concern, with those higher in empathic concern making fewer speedy decisions when choosing for others, suggesting more deliberative choice-making when choosing for the welfare of others in more empathic individuals. Trial-by-trial PEs during personal and vicarious learning were tracked in the perigenual anterior cingulate cortex (pgACC), whereas activity in the striatum specialized in coding PEs during learning for oneself. Multivariate pattern analysis showed that distinct neural patterns in the left dlPFC, TPJ, right middle temporal lobe when making decisions for the self and others and neural activity patterns in the right TPJ, dorsal ACC, and the striatum distinguished outcomes of others and self. The classification accuracy was associated with the difference in choice exploitation. Taken together, we demonstrate that both computational and neural mechanisms share commonality and distinction for personal and vicarious partner choice.

Increasing evidence suggested that the computational mechanisms of social behavior can be characterized within a RL framework (Ligneul et al. 2016; Lindström et al. 2018; Wittmann et al. 2018), such as predicting other's generosity and trustworthiness (Fareri et al. 2015; Hackel et al. 2015) and inferring the mental states of others (Rosenthal et al. 2019). Moreover, learning on behalf of others has also been characterized by RL models, such as vicarious learning to gain rewards and avoid punishment (Lockwood et al. 2016; Lindström et al. 2018; Lengersdorff et al. 2020). Here, we show that individuals learn about the cooperativeness of others through trial-and-error and that RL models can capture personal and vicarious learning mechanisms. However, individuals' choices for themselves were faster and more exploitative, and they weighed recent experiences more, as well as preferring to stay with the previously chosen cooperative partners instead of exploring new potential partners. It has been shown that people are more aversive to risk when making decisions for themselves than for others (Beisswanger et al. 2003). Thus, it is possible that individuals, when choosing partners for themselves, are less explorative and stay with previously chosen partners to avoid potential risks of new social encounters. Alternatively, when personally interacting with others, individuals may wish to avoid betraying the previous partner by choosing another partner and therefore stick with the previously chosen partner.

We further revealed agent-general and agent-specific neuro-computational mechanisms for choosing cooperative partners. First, we found that pgACC activity positively co-varied with cooperative PEs independent of beneficiaries. ACC has been linked to tracking one's own status in a social hierarchy and tracking confidence in social and nonsocial contexts (Kumaran et al. 2016; Bang and Fleming 2018). In the current task, pgACC activity tracked the differences between expected and actual cooperative (or not) behavior of the chosen partner when learning for the self

and for another agent. A separate portion of the ACC, the dorsal ACC, distinguished between outcomes for self and other in a multivariate pattern. This more dorsal ACC area has previously been linked to cooperation, competition and social learning (Chang and Sanfey 2009; Silvetti et al. 2014; Haroush and Williams 2015) but also to learning and decision-making in general (Apps et al. 2016; Kolling et al. 2016).

Second, we revealed agent-specific neural activity patterns when learning to choose cooperation partners at different stages of the learning processes. During the choice-making stage, participants showed a neural activity pattern in the TPJ and dlPFC that differentiated decisions made for another person and for oneself. Univariate responses in these areas have previously been observed when making gambling decisions for another person (Braams et al. 2014a, 2014b), suggesting the important role of the TPJ and dlPFC in vicarious decision-making (Wu et al. 2020). The TPJ and dlPFC have been implicated in mentalizing others' thoughts (Saxe 2006) and the inhibition of self-centered motives (Baumgartner et al. 2011; Buckholtz 2015). Thus, our results suggested that individuals, when choosing for others, may not simply take their first perspective to make choices; rather, they regulate their own thoughts and take the perspective of the other agent.

During the outcome-presentation stage, the TPJ and the striatum may play different roles. We found the bilateral striatal and TPJ engaged in different encoding processes, with bilateral striatum encoding PEs during self-learning and TPJ differentiating self- and other-related outcomes. Moreover, we found that individual's empathic concern was associated with striatal encoding of self (vs. other) PEs, but not with TPJ activity pattern that differentiated self and other outcomes ($r = 0.039$, $P = 0.843$). Previous studies have also linked the striatal activity to computing reward PEs when learning the association between abstract stimuli and monetary rewards (i.e. a nonsocial context, Daw et al. 2006; Lefebvre et al. 2017). Similar to the current study, other studies also reported striatal encoding of PEs in social context, such as learning about others' generosity (Hackel et al. 2015), trustworthiness (Fareri et al. 2015), and social approval (Will et al. 2017). Taken together, previous and our current findings suggested that the striatum specifically signaled self-interest-related PEs in both social and nonsocial contexts. It has also been reported that the striatum has a high response for reward to oneself than to unknown others (Albrecht et al. 2011; Braams et al. 2014a). The involvement of the striatum in encoding self-related but not other-related PE further indicates the specific role of this area may play in updating the learning experience related to oneself.

Moreover, the activity pattern of the right TPJ during the outcome stage distinguished between the self- and other-regarding outcomes. The right TPJ has been implicated in the mentalizing (Saxe 2006) and self-other distinction (Quesque and Brass 2019). Our findings thus suggested that the self-other distinction in processing outcomes might be an important feature for vicarious learning and decision-making. When making vicarious decisions, individuals may put themselves into others' shoes and consider whether other people would be satisfied with the outcome. Alternatively, the TPJ activity patterns differentiating outcomes of the self and another agent may also be associated with processing outcomes of prosocial decisions, as the TPJ has been shown to play a key role in prosociality and altruistic decision-making (Morishima et al. 2012; Hutcherson et al. 2015; Liu et al. 2019). Previous studies using monetary allocation tasks have revealed stronger TPJ activity when individuals forgo their own rewards in favor of others' benefits (Hutcherson et al. 2015; Strombach et al. 2015). Anatomical evidence also showed that the gray matter volume in the right TPJ was positively associated with individuals' altruistic behaviors (Morishima et al. 2012). Consistently, we showed that individuals showing higher classification accuracy of the right TPJ in distinguishing self and other outcomes are more exploitative to partners' cooperation when making choices for the other agent (relative to the self). Moreover, benefits for the self and for another agent were not set in conflict in the current task, thus further support the function role of the right TPJ in other-regarding motives. Since the variance of the classification performance is relatively high for small sample sizes (Combrisson and Jerbi 2015), it would be helpful to increase the sample size in the future studies to investigate the function of right TPJ in the vicarious rewarding process.

One potential limitation of our study is that it only involved personal and vicarious learning in the social context—choosing cooperation partners. To explicitly test whether the same brain regions also support the process of nonsocial learning (i.e. monetary reward learning) in self- and other-referenced frameworks, future studies could include a nonsocial condition where the agent (on behalf of self and other) would complete a task with the same structure, reward probabilities as we set here, but his/her partner is computer. In addition, it should be noted that, in the current task, participants were asked only to choose 1 of 2 players to cooperate with in each round without the opportunity to defect. It remains unknown whether individuals would also adopt a basic RL model to update their partners' choice in more complicated situations. For example, if participants play a 2-stage cooperation game where they first choose partners to play a classic PD game with and then choose to cooperate or defect, participants may employ more sophisticated learning algorithms to track others' mental states and tackle the cooperation problem (Zhu et al. 2012; Hula et al. 2018). In addition, social preference has been implicated in prosocial and cooperative choices (Fehr and Schmidt 1999; Charness and Rabin 2002; Chen and Krajbich 2018), thus the combination of the learning model and the social preference model may help us understand cooperation-partner choice in a more complex situation. It would be interesting for future studies to reveal the computational and neural mechanisms underlying different strategic cooperation on behalf of the self and other.

Overall, we reveal the computational and brain mechanisms that underpin learning about the cooperativeness of others for ourselves and on behalf of someone else. We show that common and distinct univariate and multivariate signals support this learning. These findings could have important implications for understanding how people form successful partner relationships and ultimately help to understand disorders of social learning.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) (32271092 and 32130045 to C. L.; 32125019 and 31771204 to Y. M.), the Major Project of National Social Science Foundation (19ZDA363 to C. L.), the National Key Research and Development Program of China (2022ZD0211000 to Y. M.), the Beijing Municipal Science and Technology Commission (Z151100003915122 to C. L.), and the National Program for Support of Top-notch Young Professionals to C. L.

Conflict of interest statement: The authors declare that they have no competing interests.

References

Albrecht K, Volz KG, Sutter M, Laibson DI, von Cramon DY. What is for me is not for you: brain correlates of intertemporal choice for self and other. *Soc Cogn Affect Neurosci*. 2011:6(2):218–225.

Apps MAJ, Rushworth MFS, Chang SWC. The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron*. 2016:90(4):692–707.

Aquino TG, Minxha J, Dunne S, Ross IB, Mamelak AN, Rutishauser U, et al. Value-related neuronal responses in the human amygdala during observational learning. *J Neurosci*. 2020:40(24):4761–4772.

Bang D, Fleming SM. Distinct encoding of decision confidence in human medial prefrontal cortex. *Proc Natl Acad Sci U S A*. 2018: 115(23):6082–6087.

Baumgartner T, Knoch D, Hotz P, Eisenegger C, Fehr E. Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat Neurosci*. 2011:14(11):1468–1474.

Beisswanger AH, Stone ER, Hupp JM, Allgaier L. Risk taking in relationships: differences in deciding for oneself versus for a friend. *Basic Appl Soc Psychol*. 2003:25(2):121–135.

Braams BR, Güroğlu B, de Water E, Meuwese R, Koolschijn PC, Peper JS, et al. Reward-related neural responses are dependent on the beneficiary. *Soc Cogn Affect Neurosci*. 2014a:9(7):1030–1037.

Braams BR, Peters S, Peper JS, Güroğlu B, Crone EA. Gambling for self, friends, and antagonists: differential contributions of affective and social brain regions on adolescent reward processing. *NeuroImage*. 2014b:100:281–289.

Buckholtz JW. Social norms, self-control, and the value of antisocial behavior. *Curr Opin Behav Sci*. 2015:3:122–129.

Chang LJ, Sanfey AG. Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining. *Front Behav Neurosci*. 2009:3:1–12.

Charness G, Rabin M. Understanding social preferences with simple tests. *Q J Econ*. 2002:117(3):817–869.

Chen F, Krajbich I. Biased sequential sampling underlies the effects of time pressure and delay in social decision making. *Nat Commun*. 2018:9(1):1–10.

Combrisson E, Jerbi K. Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J Neurosci Methods*. 2015:250:126–136.

Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. *J Pers Soc Psychol*. 1983:44(1): 113–126.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature*. 2006: 441(7095):876–879.

De Dreu CK, Gross J, Fariña A, Ma Y. Group cooperation, carrying-capacity stress, and intergroup conflict. *Trends Cogn Sci*. 2020:24(9):760–776.

Eldar E, Hauser TU, Dayan P, Dolan RJ. Striatal structure and function predict individual biases in learning to avoid pain. *Proc Natl Acad Sci*. 2016:113(17):4812–4817.

Fareri DS, Chang LJ, Delgado MR. Computational substrates of social value in interpersonal collaboration. *J Neurosci*. 2015:35(21): 8170–8180.

Fehr E, Fischbacher U. The nature of human altruism. *Nature*. 2003:425(6960):785–791.

Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *Q J Econ*. 1999:114(3):817–868.

Feng C, Yan X, Huang W, Han S, Ma Y. Neural representations of the multidimensional self in the cortical midline structures. *NeuroImage*. 2018:183:291–299.

Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: interactions between affect and effect. *NeuroImage*. 2012:62(1):154–166.

Hackel LM, Doll BB, Amodio DM. Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nat Neurosci*. 2015:18(9):1233–1235.

Haroush K, Williams ZM. Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell*. 2015: 160(6):1233–1245.

Harris A, Clithero JA, Hutcherson CA. Accounting for taste: a multiattribute neurocomputational model explains the neural dynamics of choices for self and others. *J Neurosci*. 2018:38(37):7952–7968.

Hebart MN, Görgen K, Haynes JD. The decoding toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinform*. 2015:8:1–18.

Hill MR, Boorman ED, Fried I. Observational learning computations in neurons of the human anterior cingulate cortex. *Nat Commun*. 2016:7(1):12722.

Hula A, Vilares I, Lohrenz T, Dayan P, Montague PR. A model of risk and mental state shifts during social interaction. *PLoS Comput Biol*. 2018:14(2):e1005935.

Hutcherson CA, Bushong B, Rangel A. A neurocomputational model of altruistic choice and its implications. *Neuron*. 2015:87(2): 451–462.

Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P. Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput Biol*. 2011:7(4):e1002028.

Kolling N, Wittmann MK, Behrens TEJ, Boorman ED, Mars RB, Rushworth MFS. Value, search, persistence and model updating in anterior cingulate cortex. *Nat Neurosci*. 2016:19(10):1280–1285.

Kumaran D, Banino A, Blundell C, Hassabis D, Dayan P. Computations underlying social hierarchy learning: distinct neural mechanisms for updating and representing self-relevant information. *Neuron*. 2016:92(5):1135–1147.

Lau T, Gershman SJ, Cikara M. Social structure learning in human anterior insula. *Elife*. 2020:9:1–17.

Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S. Behavioural and neural characterization of optimistic reinforcement learning. *Nat Hum Behav*. 2017:1(4):1–9.

Lengersdorff LL, Wagner IC, Lockwood PL, Lamm C. When implicit prosociality trumps selfishness: the neural valuation system underpins more optimal choices when learning to avoid harm to others than to oneself. *J Neurosci*. 2020:40(38):7286–7299.

Li S, Ma S, Wang D, Zhang H, Li Y, Wang J, Ma Y. Oxytocin and the punitive hub—dynamic spread of cooperation in human social networks. *J Neurosci*. 2022:42(30):5930–5943.

Ligneul R, Obeso I, Ruff CC, Dreher JC. Dynamical representation of dominance relationships in the human rostromedial prefrontal cortex. *Curr Biol*. 2016:26(23):3107–3115.

Lindström B, Haaker J, Olsson A. A common neural network differentially mediates direct and social fear learning. *NeuroImage*. 2018:167:121–129.

Liu Y, Li S, Lin W, Li W, Yan X, Wang X, Ma Y. Oxytocin modulates social value representations in the amygdala. *Nat Neurosci*. 2019:22(4):633–641.

Lockwood PL, Apps MAJ, Valton V, Viding E, Roiser JP. Neurocomputational mechanisms of prosocial learning and links to empathy. *Proc Natl Acad Sci*. 2016:113(35):9763–9768.

Lockwood PL, Wittmann MK, Apps MAJ, Klein-Flügge MC, Crockett MJ, Humphreys GW, Rushworth MFS. Neural mechanisms for learning self and other ownership. *Nat Commun*. 2018:9(1):4747.

Lockwood PL, Apps MAJ, Chang SWC. Is there a "social" brain? Implementations and algorithms. *Trends Cogn Sci xx*. 2020:24(10): 1–12.

Ma Y, Han S. Why respond faster to the self than others? An implicit positive association theory of self-advantage during implicit face recognition. *J Exp Psychol-Hum Percept Perform*. 2010:36(3):619–633.

McAuliffe WHB, Burton-Chellew MN, McCullough ME. Cooperation and learning in unfamiliar situations. *Curr Dir Psychol Sci*. 2019:28(5):436–440.

Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biol Psychiatry*. 2017:82(7):532–539.

Morelli SA, Sacchet MD, Zaki J. Common and distinct neural correlates of personal and vicarious reward: a quantitative meta-analysis. *NeuroImage*. 2015:112:244–253.

Morishima Y, Schunk D, Bruhin A, Ruff CC, Fehr E. Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*. 2012:75(1):73–79.

Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. *NeuroImage*. 2005:25(3): 653–660.

Noë R, Hammerstein P. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav Ecol Sociobiol*. 1994:35(1):1–11.

Quesque F, Brass M. The role of the temporoparietal junction in self-other distinction. *Brain Topogr*. 2019:32(6):943–955.

Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD. A neural basis for social cooperation. *Neuron*. 2002:35(2):395–405.

Rosenthal IA, Hutcherson CA, Adolphs R, Stanley DA. Deconstructing theory-of-mind impairment in high-functioning adults with autism. *Curr Biol*. 2019:29(3):513–519.e6.

Saxe R. Uniquely human social cognition. *Curr Opin Neurobiol*. 2006: 16(2):235–239.

Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R. Serotonin selectively modulates reward value in human decision-making. *J Neurosci*. 2012:32(17):5833–5842.

Silvetti M, Alexander W, Verguts T, Brown JW. From conflict management to reward-based decision making: actors and critics in primate medial frontal cortex. *Neurosci Biobehav Rev*. 2014:46: 44–57.

Singer T, Seymour B, Kaube H, Dolan RJ, Frith CD. Empathy for pain involves the affective but not sensory components of pain. *Science (80-)*. 2004:341(5661):1157–1162.

Stallen M, Sanfey AG. The cooperative brain. *Neuroscience*. 2013:19: 292–303.

Stirrat M, Perrett DI. Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychol Sci*. 2010:21(3): 349–354.

Strombach T, Weber B, Hangebrauk Z, Kenning P, Karipidis II, Tobler PN, Kalenscher T. Social discounting involves modulation of neural value signals by temporoparietal junction. *Proc Natl Acad Sci U S A*. 2015:112(5):1619–1624.

Sul S, Tobler PN, Hein G, Leiberg S, Jung D, Fehr E, Kim H. Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proc Natl Acad Sci U S A*. 2015:112(25):7851–7856.

Takahashi C, Yamagishi T, Tanida S, Kiyonari T, Kanazawa S. Attractiveness and cooperation in social exchange. *Evol Psychol*. 2006:4(1):147470490600400.

Will GJ, Rutledge RB, Moutoussis M, Dolan RJ. Neural and computational processes underlying dynamic changes in self-esteem. *Elife*. 2017:6:e28098.

Wittmann MK, Lockwood PL, Rushworth MFS. Neural mechanisms of social cognition in primates. *Annu Rev Neurosci*. 2018:41(1): 99–118.

Wu H, Liu X, Hagan CC, Mobbs D. Mentalizing during social InterAction: a four component model. *Cortex*. 2020:126:242–252.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*. 2011:8(8):665–670.

Zhu L, Mathewson KE, Hsu M. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proc Natl Acad Sci U S A*. 2012:109(5):1419–1424.

Zhu R, Feng C, Zhang S, Mai X, Liu C. Differentiating guilt and shame in an interpersonal context with univariate activation and multivariate pattern analyses. *NeuroImage*. 2019:186:476–486.